

20
24



UNIVERSIDAD NACIONAL AUTÓNOMA
DE MÉXICO

FACULTAD DE CIENCIAS

ESTUDIOS DE ESTABILIDAD Y ESTRUCTURA
ELECTRÓNICA EN MOLÉCULAS ADN-TRIPLEX

T E S I S

QUE PARA OBTENER EL TÍTULO DE

FÍSICO

PRESENTA:

ADRIÁN ARTURO HUERTA HERNÁNDEZ



FACULTAD DE CIENCIAS
UNAM

**TESIS CON
FALLA DE ORIGEN**



FACULTAD DE CIENCIAS
SECCION ESCOLAR

1996

**TESIS CON
FALLA DE ORIGEN**



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.



UNIVERSIDAD NACIONAL
AUTÓNOMA DE
MÉXICO

M. en C. Virginia Abrín Batule
Jefe de la División de Estudios Profesionales de la
Facultad de Ciencias
P r e s e n t e

Comunicamos a usted que hemos revisado el trabajo de Tesis: "Estudios de estabilidad y estructura electrónica en moléculas ADN-triplex"

realizado por Adrián Arturo Huerta Hernández

con número de cuenta 8615581-2 , pasante de la carrera de Física

Dicho trabajo cuenta con nuestro voto aprobatorio.

Atentamente

Director de Tesis
Propietario

DR. JUAN ANTONIO COGORDAN RAMIREZ

Propietario

DR. ALIPIO GUSTAVO CALLES MARTINEZ

Propietario

DR. GERMINAL COCHO GIL

Suplente

DR. IVAN ORTEGA BLAKE

Suplente

DR. FRANCISCO XAVIER SOBERON MAINERO

Consejo Departamental de Física

DR. ROBERTO ALEJANDRO RUELAS MAYORGA
Coordinador de Licenciatura

*A mis padres, familiares y amigos.
(Gracias por su apoyo y confianza)*

Contenido

1	Introducción	6
I	Antecedentes	9
2	Estructura molecular del ADN	10
2.1	Estructura primaria	10
2.2	Estructura secundaria; <i>duplex</i> y <i>triplex</i>	11
3	Modelado molecular	17
3.1	Física molecular	17
3.1.1	Aproximación cuántica (Born-Oppenheimer)	18
3.2	Problema electrónico	20
3.3	Problema nuclear, aproximación para sistemas macromoleculares	21
3.3.1	Campo de fuerza para moléculas biológicas, AMBER	23
II	Metodología y Desarrollo	26
4	Simulación computacional	27
4.1	Construcción del sistema	27
4.1.1	Cálculo de cargas atómicas para C^+	29
4.1.2	Posición de contraiones	30
4.1.3	Configuración inicial	32
4.1.4	Simulación con dinámica molecular	33
4.2	Análisis e interpretación de resultados	35
4.2.1	Parámetros intermoleculares	37
4.2.2	Estructura secundaria del ADN- <i>triplex</i>	38
4.2.3	Estructura electrónica molecular	46
4.2.4	Estructura del Solvente	51
5	Conclusiones	60

	3
6 Apendices	63
6.1 Método de Hartee-Fock	63
6.2 Orden de enlace	65
6.3 Análisis de población de Mulliken	66
6.4 Superficie molecular de potencial electrostático	67
6.5 Método de dinámica molecular	67
6.5.1 Ensamblés NVE, NVT y NPT.	69
6.6 Programa para calcular la posición de contraiones	70
Bibliografía	75

Agradecimientos

Principalmente deseo agradecer a la UNAM por el alto nivel de conocimientos y cultura que pone en nuestras manos, así también a la Facultad de Ciencias cuya magnífica planta docente que en ella labora merece mi respeto y reconocimiento.

Deseo agradecer particularmente al Dr. Juan Antonio Cogordan, quien apartir del curso de Física Moderna III de la carrera, me despertó el interés por el uso de sistemas computacionales en la Física. Al Dr. Mayo Villagran, quien durante mi Servicio Social me encamino sobre las aplicaciones de la Física en sistemas de interés biomédico. Al Dr. Jaime Lagunez por el material e información sobre sistemas los ADN-*triplex* y sus implicaciones biomédicas. Al Dr. Orest Pizío y Dr. Andrei Trokhymchuck por sus comentarios y discusiones.

También extendo mi agradecimiento al Dr. Alipio Calles, Dr. Germinal Cocho, Dr. Ivan Ortega y Dr. Francisco Xavier Soberon por la revisión y las correcciones hechas al texto, las cuales contribuyeron a mejorarlo.

A mis compañeros, que durante toda la carrera me han brindado su amistad: Ernesto Bautista, Alejandro Camberos, Gabriela Carreon, Armando Flores, Arcadio Huerta, Margarita Mayoral, Roman Linares, Luis Perez y muy especialmente a Claudia Jacal.

A la DGSCA-UNAM por facilitarme sus excelentes recursos en supercomputo, visualización e impresión y de atención a usuarios.

Al Instituto de Química por permitirme el uso de sus instalaciones.

Agradezco los apoyos económicos recibidos de la DGAPA y de los fondos para investigación por parte del convenio UNAM-CRAY.

Prologo

El conocimiento científico aplicado con múltiples objetivos ha dado como producto el impresionante desarrollo tecnológico que caracteriza al presente siglo, en donde los sistemas computacionales han influido de manera notoria. La computación se ha convertido en una valiosa herramienta para diversas áreas de estudio; por un lado facilita el cálculo y apoya con utilerías gráficas la interpretación de resultados. Por otra parte, es también eficiente en el control, el procesamiento de datos y las telecomunicaciones, con los que se agiliza el manejo y obtención de información. En particular, la física contemporánea hace uso del cómputo en múltiples actividades tanto teoricas como experimentales. A partir de los cursos que tome en la Facultad de Ciencias, en los que se ha enfatizado el cómputo como una herramienta útil para la Física y otras áreas afines, me surgió el interés por el desarrollo de la presente tesis. Esto motivado, como mencionaré a lo largo del texto, por el apoyo que las aplicaciones de la Física y el cómputo dan en areas como la Química, Biología e incluso la Medicina.

La tesis consta de una introducción que ubica al lector en el tema de estudio y da un panorama general de su interés biomédico. Posteriormente los siguientes capítulos se dividen en dos partes; la primera de **antecedentes** donde menciono aspectos importantes que me sirvieron para conocer algunas generalidades sobre la estructura del ADN y los modelos moleculares ya existentes. En la segunda parte, describo la **metodología** usada, **el desarrollo, el análisis y los resultados** obtenidos del presente estudio, junto con el análisis hago una pequeña discusión de cada uno de los puntos que conforman mi interpretación de los resultados y por último, el capítulo de las **conclusiones** que hago sobre el desarrollo de todo el trabajo. Debo mencionar que aún cuando las interpretaciones que doy me producen ciertas dudas e inquietudes en aspectos relacionados con la Biología, Química y Medicina, las presento tal cual las he pensado.

Capítulo 1

Introducción

El tema de la presente tesis es de interés para la Medicina dada la necesidad de obtener fármacos para atacar enfermedades. Actualmente se sabe que algunas proteínas están involucradas o favorecen distintos tipos de enfermedades, con esto en mente se ha buscado diseñar drogas capaces de llegar e interactuar en sitios específicos de estas proteínas para inhibir su actividad. Recientemente se ha tornado la atención por otro tipo de fármacos con la propiedad de seleccionar sitios específicos de los ácidos nucleicos (*ADN* y *ARN*) e inhibir la síntesis de dichas proteínas, peculiaridad que se ha encontrado en pequeños fragmentos de cadenas de ADN llamados *oligos*, a los que me referiré posteriormente. Aprovechando este hecho se ha trabajado en dos estrategias que impiden la síntesis de proteínas a diferentes niveles, llamadas *antigene* y *antisense*. Es decir, tomando en cuenta que el mecanismo para la síntesis de proteínas consiste; en primer lugar en **transcribir** el código de un gen¹ en el ADN a un ARN mensajero (ARNm) y posteriormente **traducir** dicha información del ARNm a la proteína (fig. 1.1). La estrategia llamada *antigene* bloquea la transcripción por la adición de un oligo a la estructura del ADN, formándose una estructura triple (*ADN-triplex*) que no es reconocida por la ARN polimerasa, enzima que transcribe el código al ARNm, inhibiendo de esta forma la síntesis de la proteína. La *antisense*, de forma semejante bloquea la traducción añadiendo para ello un oligo al ARNm y de esta forma el ribosoma no sintetiza la proteína. Con estas estrategias se ha logrado disminuir los efectos *in vitro* e *in vivo* de algunos virus y otros tipos de enfermedades [Cohen, et al, 1994].

No obstante, existen limitaciones que disminuyen su efectividad [Stein, et al, 1993], de donde surgen interrogantes como las siguientes: ¿Cómo sintetizar fácilmente los oligos?;

¹Un gen contiene la información necesaria para sintetizar la proteína.

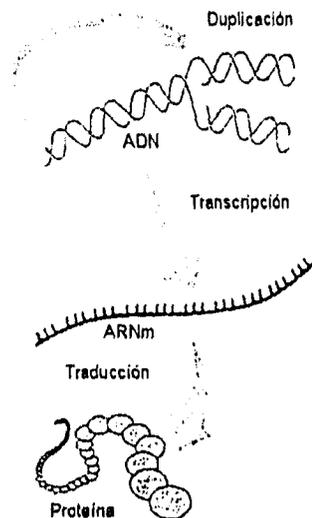


Figura 1.1: Esquema que representa diferentes fases en el flujo de información genética.

¿Cómo hacer que los oligos entren y se retengan en el interior de las células?; ¿Cómo hacer para que las estructuras *triplex* se formen y se mantengan estables *in vivo*? ó ¿Cómo impedir que el oligo interactue con otras zonas del ADN y ARNm o con otras macromoléculas?. Respuestas a todas estas interrogantes son difíciles de proporcionar, sin embargo pueden provenir desde muy diversas fuentes.

A partir de que en 1940 se identificara el ácido desoxirribonucleico (ADN) como material genético, entre los años 50 y 60 se desencadenaron una serie de estudios relacionados tanto con su estructura como con la forma en que este codifica y transmite la información hereditaria, preguntas tales como: ¿Cómo hace el ADN para reproducirse y transmitir la información genética de generación en generación?; ¿Cuál es el mecanismo para la síntesis de proteínas y que sistema es usado para relacionar una secuencia de nucleótidos del ADN con una secuencia de aminoácidos de una proteína?, no podían ser contestadas hasta que en 1953 James D. Watson y Francis Crick [Watson-Crick, et al, 1953] lograron conjuntar toda la información obtenida al respecto sobre las propiedades biofísicas y fisicoquímicas, para así proponer la **estructura secundaria** de doble hélice (*duplex*) del ADN. De la misma manera, considero que el mejor conocimiento de estas propiedades pueden proporcionar guías o pequeñas soluciones para aumentar la efectividad de las estrategias. Con esto

Parte I
Antecedentes

Capítulo 2

Estructura molecular del ADN

Para familiarizarme con algunas de las propiedades físicas del ADN y poder desarrollar el presente estudio, fue necesario hacer un repaso sobre su estructura molecular. En el presente capítulo pretendo exponer algunos de los aspectos estructurales del ADN, así como los factores que se sabe están involucrados con su estabilidad, introduciendo algunas de las definiciones con la nomenclatura comúnmente usada en este tipo de sistemas.

Los genes que controlan las características hereditarias y la síntesis de las proteínas son moléculas de ácido nucleico; desoxirribonucleico (ADN) y ribonucleico (ARN). Estos ácidos nucleicos, al igual que las proteínas son macromoléculas que constituyen la materia viviente. Una característica general de estos sistemas macromoleculares es la formación de diferentes estructuras, a las que me referiré a continuación.

2.1 Estructura primaria

Con algunos datos fisicoquímicos (e.g. sedimentación, difusión, dispersión de luz y medidas de viscosidad [Watson-Crick, et al, 1953]) junto con la observación de las estructuras cristalinas, que inició Furberg en 1951 [Saenger], se pudo deducir que los ácidos nucleicos son fibras largas y delgadas que están formadas con unidades monoméricas llamadas **nucleótidos**. Estas fibras son en realidad heteropolímeros, puesto que están constituidas por cuatro nucleótidos diferentes. Cada nucleótido contiene tres componentes; un **azúcar** (desoxirribosa), un grupo de **fosfato** (fosfodiéster) y una **base** de nitrógeno. En el ADN las bases son **Timina Citosina Adenina y Guanina**, donde las dos primeras bases son monocíclicas **pirimidinas** y las dos últimas bicíclicas **purinas** (fig. 2.1). El azúcar y el

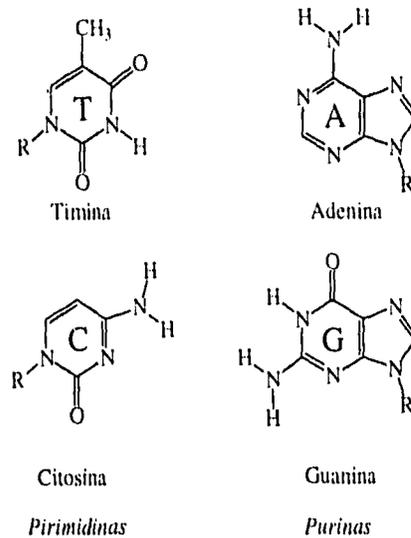


Figura 2.1: Bases de nitrógeno del ADN. La R en las moléculas es el lugar donde se une al azúcar.

fosfato se van alternando a lo largo de la fibra formando una cadena continua, unidos en los carbonos del azúcar en los sitios llamados 3' y 5'. Cada azúcar tiene unida una base de nitrógeno que sale de este (fig. 2.2), esta cadena es llamada **estructura primaria** de las moléculas de ADN. El ARN es similar a una cadena del ADN, excepto que el azúcar es una ribosa en lugar de desoxirribosa y tiene una base de **Uracilo** en lugar de timina.

2.2 Estructura secundaria; *duplex* y *triplex*

La forma en la que el azúcar y los fosfatos se unen ocasiona que el ADN tenga dos extremos diferentes, caracterizados por las posiciones 3' y 5' antes mencionadas¹. La **estructura secundaria** del ADN más común, esta formada por cadenas antiparalelas (3'-5') y (5'-3') en forma de *duplex*, en donde las cadenas se encuentran ligadas mediante **puentes de hidrógeno** entre cada una de sus bases. Por requerimientos estéricos es más probable² formar estos puentes entre T:A y C:G, por lo que cada cadena es complemento de

¹Esto ayuda a que la transcripción del código genético sólo pueda hacerse en una dirección (5'→3') [Mathews].

²Eventualmente ocurre la formación de los llamados *mismatch* en los cuales se forman puentes de hidrógeno más débiles (e.g. A:A, G:G, C:C⁺, T:G, [Mergny, et al, 1991]), esto puede provocar que el oligo se

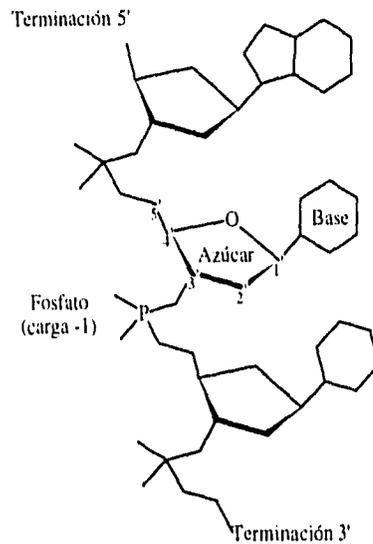


Figura 2.2: Esquema en el que se muestran los grupos principales que conforman la estructura primaria del ADN.

la otra. Al **par de bases** enlazadas de esta forma se les conoce con el nombre de **Watson-Crick** existiendo también la posibilidad de formar puentes de hidrógeno entre las bases en orden inverso A:T conocido con el nombre de **Hoogsteen** (fig. 2.3). Motivado por fuerzas de dispersión y efectos hidrofóbicos, las bases en el ADN-*duplex*, se apilan a lo largo de las cadenas formándose dos surcos, conocidos como *major groove* y *minor groove* (fig. 2.5). Características estructurales (e.g. ángulos formados entre enlaces, distancias entre bases y anchos del *major* y *minor groove*) determinan el **polimorfismo estructural**³ del ADN, esto se ha obtenido mediante estudios experimentales hechos con difracción de rayos X en fibras de ADN, catalogándose en tres grupos representativos llamados A-ADN, B-ADN y Z-ADN, en donde cada forma depende de la humedad relativa, la concentración iónica del ambiente y de la propia secuencia, ([Saenger] p. 253). Para las estructuras secundarias de ADN-*triplez* aún no existen suficientes estudios de esta naturaleza [Laughton, et al, 1992].

La estructura secundaria ADN-*triplez* mejor conocida es formada con **homopolynucleótidos**⁴, ocurre debido a que en el *major groove* de una ADN-*duplex* homopirini-

una erróneamente en secuencias muy semejantes a la cual se dirija.

³Capacidad que tiene la estructura de adoptar diversas formas.

⁴Cuando una cadena o fragmento de ella, está formada solamente de purinas es llamada homopurina,

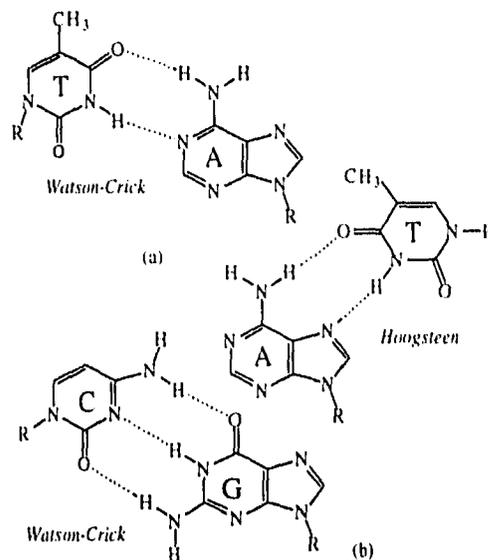


Figura 2.3: Unión par de las bases. a) T:A (Watson-Crick) y A:T (Hoogsteen), b) C:G (Watson-Crick).

dina/homopurina, enlazada con puentes *Watson-Crick*, hay el espacio suficiente para estabilizar mediante puentes de hidrógeno tipo *Hoogsteen* un oligodeoxynucleótido (**oligo**⁵) homopirimidina, en orientación paralela con respecto a la cadena homopurina de la *duplex*. Esta estructura fue observada por primera vez en 1957 [Felsenfeld, et al, 1957]. En ocasiones sucede naturalmente, que una cadena homopirimidina de la propia doble hélice se dobla y se enrolla su *major groove*, dando origen a una estructura *triple* conocida como H-ADN, que se cree juega un papel importante en la autorregulación de la expresión genética [Laughton, et al, 1992]. No obstante en este tipo de formación natural se encuentra otra limitante, debido a que sólo se forma con la **unión triple de bases T:A:T** y no da lugar a la formación de secuencias alternantes (fig. 2.4 (a)). Sobre estructuras formadas con las uniones T:A:T se cuenta con información obtenida por difracción de fibra [Arnott, et al, 1974-76], clasificándose su forma en un A-ADN.

Con el fin de ampliar la posibilidad de las estrategias se han buscado nucleótidos de la misma forma, si esta formada con pirimidinas son llamadas **homopirimidinas**.

⁵Se le llama oligo, por la poca cantidad de nucleótidos (decenas), comparado con los que tiene una cadena del ADN (dependiendo del organismo, cientos, miles, millones, etc.).

capaces de establecer puentes de hidrógeno entre el oligo y el ADN-*duplex*, una forma de lograr esto es mediante la **protonación**⁶ de los nucleótidos C del oligo creandose la unión triple de bases C:G:C⁺ con puentes de hidrógeno tipo *Hoogsteen* (fig. 2.4 (b)). No obstante, aún cuando la protonación del nucleótido ocurre con el aumento de acidez del medio, lo cual disminuye las posibilidades de la estrategia *antigene* [Laughton, et al, 1992], [Dervan, et al, 1989] y [Radhakrishnan, et al, 1991], consideramos que iniciar con el estudio de este tipo de sistemas nos proporcionará la información necesaria para continuar con la búsqueda de estructuras más estables, dada la importancia de sus propiedades.

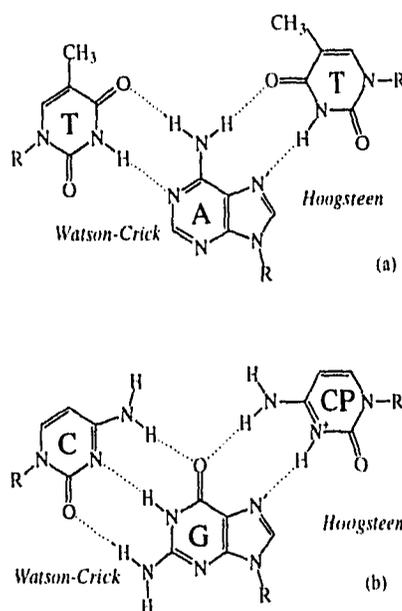
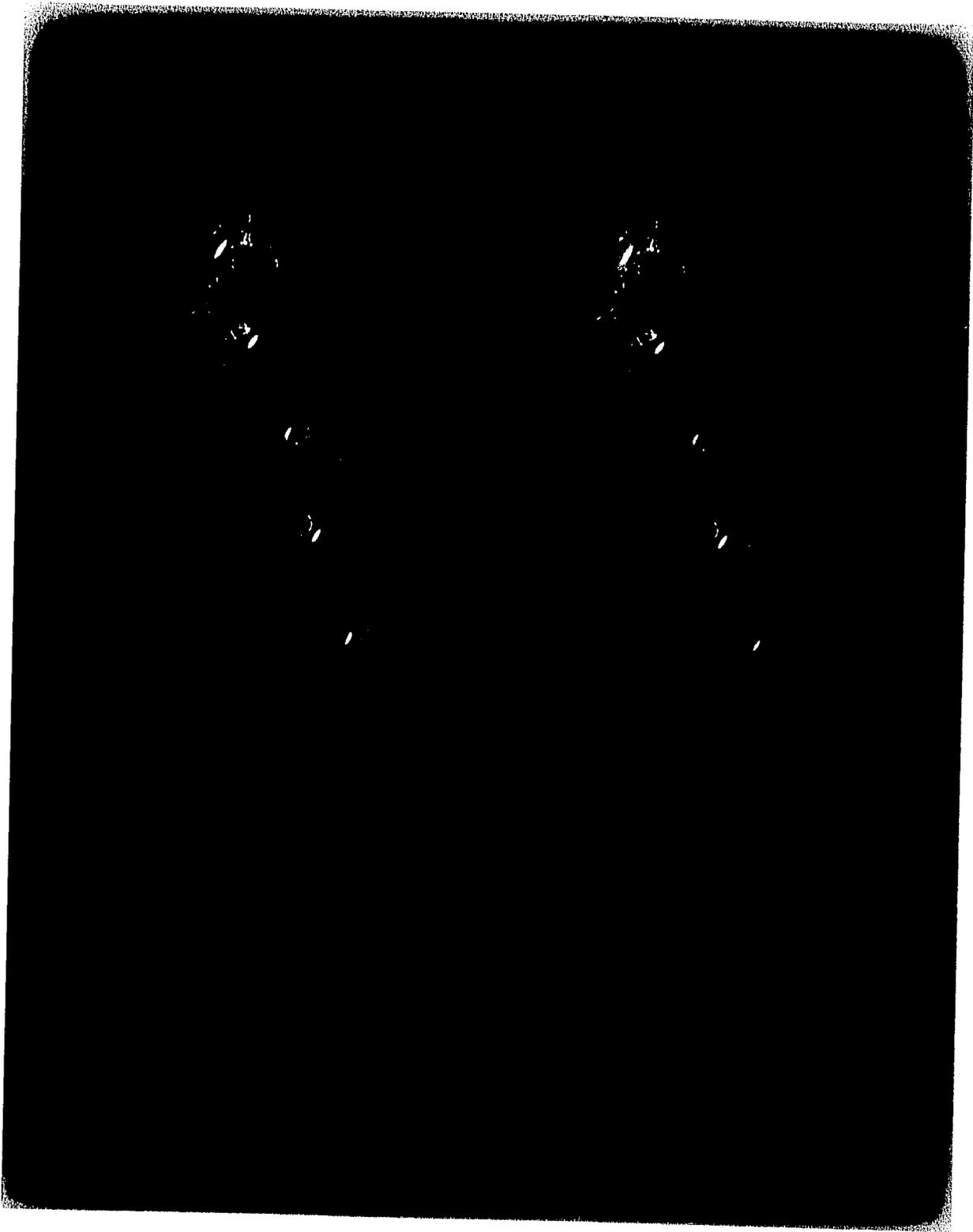


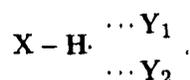
Figura 2.4: Unión triple de las bases. a) T:A:T y b) C:G:C⁺, mostrando los puentes de hidrógeno formados.

Variaciones de temperatura o cambios de acidez (pH) del medio [Dervan, et al, 1989] pueden provocar la pérdida de la estructura secundaria del ADN en cadenas simples, a este proceso se le llama **desnaturalización**. La naturaleza electrolítica del ADN origina una fuerte repulsión electrostática entre las cadenas, esto es debido a que el ADN al estar en solución a pH fisiológico su grupo fosfato se ioniza formándose aniones a lo largo de las

⁶La protonación consiste en agregar un protón (núcleo del hidrógeno) al nucleótido.



cadenas (polianiones), no obstante gracias a la presencia en el medio de **contraiones**⁷, de la formación de los puentes de hidrógeno, así como de la interacción con el solvente del medio, las estructuras secundarias del ADN se mantienen estables bajo condiciones fisiológicas y no se desnaturalizan. De aquí que los factores más importantes que favorecen la estabilidad de la estructura secundaria de las moléculas de ADN, sean los **puentes de hidrógeno**, las **fuerzas de Van der Waals** y los **efectos hidrofóbicos**. Como ya se mencionó los puentes de hidrógeno unen las bases entre las cadenas⁸, en general un puente de hidrógeno será representado $X - H \cdots Y$ y es debido a que un átomo de hidrógeno H conecta dos átomos X y Y de gran electronegatividad. Por esto último tiene un carácter principalmente electrostático y su intensidad depende de las cargas atómicas de X, H y Y, esto se refleja en la separación entre $H \cdots Y$ (normalmente entre 2-3 Å). Al grupo $X - H$ se le conoce como **grupo donador** y a Y como **aceptor**. A diferencia de los enlaces covalentes, los puentes de hidrógeno son de 20 a 30 veces más débiles y poco direccionales. Incluso puede formarse cooperación entre 2 ó más receptores.



Las interacciones de puentes de hidrógeno entre las bases del ADN son del tipo $N - H \cdots N$ y $N - H \cdots O$, [Saenger].

⁷Iones de signo contrario, en este caso cationes (como K^+ , Na^+ , Mg^{2+} , Ca^{2+} y Li^+).

⁸Un punto que es interesante mencionar en este aspecto, del cual no me ocupare en este momento, es la proposición de Lödwin acerca de que las mutaciones que suceden de forma natural pueden ser debidas al efecto túnel de uno ó más protones de los puentes de hidrógeno del ADN de una base sobre una cadena a la base correspondiente sobre la otra cadena ([Levine], 2a. ed., 1974).

Capítulo 3

Modelado molecular

La descripción teórica de un fenómeno físico usualmente se basa en un **modelo** que haga abstracción de este, a partir de las observaciones que se han hecho de él en la naturaleza. Para ser aceptado físicamente, el modelo debe basarse en alguna teoría y comúnmente tener una representación matemática. Una simulación molecular puede recurrir a modelos moleculares basados en teorías, desde los mecánico-cuántico-relativistas hasta las clásicas, donde la utilidad de cada uno depende de la naturaleza del problema. Al describir sistemas macromoleculares biológicos, regularmente se recurre a una aproximación clásica.

En este capítulo hago un corto repaso tanto de los modelos cuánticos como modelos clásicos que usé durante el desarrollo de la simulación, lo cual también fue útil para dar una interpretación de los resultados obtenidos. Este repaso se consideró conveniente para iniciar el trabajo con antecedentes de algunos puntos importantes de los modelos y métodos usados.

3.1 Física molecular

Para hacer un estudio de un sistema molecular como el que ahora nos interesa, no contamos con una teoría molecular completamente satisfactoria que incluya fenómenos cuántico-relativistas ([Cohen-T], p. 9). No obstante, aún cuando los efectos relativistas añaden correcciones a sistemas tan simples como los átomos hidrogenoides ([Peña], p. 711), una aproximación cuántica es suficiente para describir un gran número de sistemas incluyendo los biológicos, por lo que la mecánica cuántica no relativista se considera un buen

punto de partida [Weissbluth].

3.1.1 Aproximación cuántica (Born-Oppenheimer)

Una aproximación cuántica implica que debemos encontrar la solución de la ecuación de Schrödinger,

$$H\Psi(\mathbf{r}, \mathbf{R}) = E\Psi(\mathbf{r}, \mathbf{R}). \quad (3.1)$$

La cual se ha considerado independiente del tiempo por simplicidad. En donde H es el operador hamiltoniano de un sistema de electrones y núcleos descritos por los vectores de posición \mathbf{r} y \mathbf{R} respectivamente. Un hamiltoniano que contiene las interacciones más importantes para una molécula de n -electrones en ausencia de campos externos, es:

$$H = H_0 + H_1 + H_2,$$

con:

$$\begin{aligned} H_0 &= - \sum_{i=1}^N \frac{1}{2} \nabla_i^2 - \sum_{A=1}^M \frac{1}{2M_A} \nabla_A^2 - \sum_{i=1}^N \sum_{A=1}^M \frac{Z_A}{r_{iA}}, \\ H_1 &= \sum_{i=1}^N \sum_{j>i}^N \frac{1}{r_{ij}} + \sum_{A=1}^M \sum_{B>A}^M \frac{Z_A Z_B}{R_{AB}}, \\ H_2 &= \sum_{i=1}^N \xi(r_i) \mathbf{l}_i \cdot \mathbf{s}_i. \end{aligned}$$

Donde r_{iA} es la distancia del núcleo A al i -ésimo electrón, r_{ij} es la distancia entre los electrones i -ésimo y j -ésimo, M_A la razón de la masa del núcleo A a la masa de un electrón y Z_A el número atómico del núcleo A . El término H_0 contiene la energía cinética y potencial de los núcleos y electrones, H_1 representa la repulsión coulombiana entre pares de electrones y de núcleos; H_2 es la interacción espín-órbita, que bajo ciertas consideraciones de simetría es posible incluir en la descripción molecular ([Weissbluth], p. 561).

Las contribuciones debidas a H_1 y H_2 dependen del número atómico, para Z pequeño es dominante H_1 comparado con H_2 y contrariamente para Z mayores H_2 es dominante comparado con H_1 . En los casos límite tendremos los siguientes Hamiltonianos.

$$H_I = H_0 + H_1,$$

$$H_{II} = H_0 + H_2.$$

Debido a que en general los sistemas biológicos están constituidos de átomos ligeros, Z pequeños, es razonable tomar el primer límite H_I para representar el Hamiltoniano total del sistema.

Por lo cual el Hamiltoniano total que consideraremos para N electrones y M núcleos es en unidades atómicas,

$$H = H_I = - \sum_{i=1}^N \frac{1}{2} \nabla_i^2 - \sum_{A=1}^M \frac{1}{2M_A} \nabla_A^2 - \sum_{i=1}^N \sum_{A=1}^M \frac{Z_A}{r_{iA}} + \sum_{i=1}^N \sum_{j>i}^N \frac{1}{r_{ij}} + \sum_{A=1}^M \sum_{B>A}^M \frac{Z_A Z_B}{R_{AB}}. \quad (3.2)$$

Para resolver la ecuación (3.1) con el Hamiltoniano (3.2) de un sistema molecular, se recurre a una aproximación conocida con el nombre de **Born-Oppenheimer**. Una discusión cualitativa de esta aproximación se basa en el siguiente hecho:

'Los electrones por moverse mucho más rápido que los núcleos atómicos, al ser estos últimos más pesados, es posible desacoplar su movimiento del movimiento nuclear.'

Es decir, podemos separar la ecuación (3.1), en una ecuación para el **problema electrónico** y otra para el **problema nuclear**, mediante el siguiente procedimiento:

1. Consideraremos que los electrones se mueven bajo el campo producido por los núcleos fijos con coordenadas R' , por lo que en el Hamiltoniano (3.2), se puede despreciar el término de energía cinética del movimiento nuclear y el término de repulsión nuclear será considerado como una constante. Definiéndose el Hamiltoniano electrónico como:

$$H_e = - \sum_{i=1}^N \frac{1}{2} \nabla_i^2 - \sum_{i=1}^N \sum_{A=1}^M \frac{Z_A}{r_{iA}} + \sum_{i=1}^N \sum_{j>i}^N \frac{1}{r_{ij}} + \underbrace{\sum_{A=1}^M \sum_{B>A}^M \frac{Z_A Z_B}{R'_{AB}}}_{\text{Constante}}. \quad (3.3)$$

2. Rescribiendo la ecuación de Schrödinger bajo la suposición anterior, se obtiene:

$$H_e \Psi_e(r; R') = E_e(R') \Psi_e(r; R'). \quad (3.4)$$

Donde la función $\Psi_e(r; R')$ es una función de onda electrónica la cual depende paramétricamente de las posiciones nucleares y $E_e(R')$ es el eigenvalor de energía correspondiente a cada configuración nuclear R' .

3. Proponiendo a la función de onda como el producto de una función de onda electrónica y una función de onda nuclear,

$$\Psi(r, R) = \Psi_e(r; R) \Psi_n(R). \quad (3.5)$$

De la ecuación (3.1) y (3.4) se obtiene que

$$H_n \Psi_n(R) = E \Psi_n(R), \quad (3.6)$$

definiendo a:

$$H_n = - \sum_{A=1}^M \frac{1}{2M_A} \nabla_A^2 + E_e. \quad (3.7)$$

El significado físico de esta aproximación es que la función de onda nuclear de la ecuación (3.6) es calculada debido al **potencial efectivo electrónico** E_e , el cual es obtenido a partir de la ecuación (3.4) para cada configuración nuclear.

Bajo este esquema ahora nos ocuparemos de resolver las ecuaciones (3.4) como el **problema electrónico** y 3.6 como el **problema nuclear** por separado.

3.2 Problema electrónico

La búsqueda de la solución por diferentes métodos de la ecuación (3.4) ha dado lugar a una gran cantidad de códigos de programación debido a su complejidad. Estos códigos en general se agrupan en dos tipos de cálculos moleculares conocidos con los nombres de *ab initio* y **semiempíricos**. Los métodos semiempíricos usan un hamiltoniano más simple que el hamiltoniano molecular correcto y usan parámetros cuyos valores son ajustados con datos experimentales o resultados de cálculos *ab initio*, ([Levine], p.455). En contraste los métodos *ab initio*¹, generalmente usan el hamiltoniano electrónico de la misma forma en que se encuentra en la ecuación (3.3). En general ambos métodos han demostrado ser muy útiles, sin embargo la elección del método más apropiado depende tanto de la cantidad de átomos que tenga el sistema molecular, del número atómico de cada uno de ellos y en el caso de los métodos semiempíricos del tipo de sistema que se esté estudiando.

En el apéndice 6.3 se muestra el método de Hartee-Fock, en él puede observarse que para encontrar las funciones de onda necesarias para resolver el problema electrónico, se tienen que evaluar un considerable número de integrales. Lo cual ocasiona que el tiempo de cómputo necesario para llevarlas al cabo se incremente conforme al número de electrones que tenga el sistema. En general esto ocurre en los métodos *ab initio*, por lo que para ahorrar tiempo se recurre a los métodos semiempíricos. Los métodos semiempíricos que

¹El nombre *ab initio*, proviene de la misma palabra en latín que significa "desde el principio" e indica que los cálculos son basados en los principios fundamentales. *Ibidem*.

comúnmente son usados en sistemas biomoleculares como los que estamos estudiando son MNDO (*Modified neglect of diatomic overlap*), AM1 (*Austin model 1*) y PM3, por representar adecuadamente sus propiedades moleculares ya que sus parámetros son ajustados con datos de experimentos en moléculas de este tipo ([Levine], p. 448).

3.3 Problema nuclear, aproximación para sistemas macro-moleculares

A partir de la separación hecha en la sección 3.1.1, con la aproximación Born-Oppenheimer se encuentra que la ecuación que describe el movimiento nuclear es (3.6). Sin embargo, calcular el potencial efectivo electrónico E_e implica un gran esfuerzo computacional para una sola configuración nuclear R , como se ha visto en la sección 3.2. En moléculas diatómicas, triatómicas o sistemas periódicos, es razonable calcular $E_e(R_i)$ para distintas configuraciones R_i y con esta información ajustar un potencial analítico $V(R)$ tal que $V(R_i) = E_e(R_i)$, obteniendo de esta manera una hipersuperficie de energía potencial con la que sea posible resolver la ecuación (3.6). El procedimiento para resolver esta ecuación es conocido como **dinámica cuántica** ya que para encontrar la evolución temporal de la molécula, es decir las diferentes configuraciones dependientes del tiempo, basta multiplicar a la función de onda Ψ_n del problema nuclear por el término temporal adecuado.

Desafortunadamente cálculos como los anteriores sólo son posibles para moléculas pequeñas y en la actualidad para sistemas de estado sólido². No siendo así para sistemas macromoleculares biológicos como el que ahora nos interesa, por su gran cantidad de átomos y su falta de simetría. No obstante, el potencial analítico $V(R)$ que genera el **campo de fuerzas** que siente cada núcleo puede ser utilizado para resolver nuestro problema macromolecular, haciendo para ello una aproximación clásica del modelo cuántico con ayuda del teorema de Hellman-Feynmann³ de la siguiente manera: *Por ser los núcleos atómicos objetos relativamente pesados*, despreciaremos sus efectos cuánticos, aproximando la ecuación (3.6) por las ecuaciones de movimiento de Newton ecuación (3.8), de manera análoga a la cuántica este procedimiento es llamado **dinámica molecular (DM)**.

²Con algunas modificaciones al esquema que estoy planteando, ha sido posible usando los algoritmos de Car y Parrinello, R.Car & M.Parrinello, Phys. Rev. Lett, vol. 55, 2471, 1985.

³El Teorema de Hellman-Feynmann (1937,1938), nos dice que mediante la aproximación de Born-Oppenheimer las fuerzas que actúan sobre un núcleo son derivables del potencial clásico $V(R)$ y la función de onda electrónica. Weissbluth, *op. cit.*, p. 555.

$$-\frac{dV(R)}{dR} = m \frac{d^2 R}{dt^2}. \quad (3.8)$$

En el potencial analítico $V(R)$ es conveniente incluir términos relacionados con los átomos unidos con enlaces $V_{ligados}$ y los no-ligados $V_{no-ligados}$. De tal forma que,

$$V(R) = V_{ligados} + V_{no-ligados}.$$

Para ajustar cada término se toman fragmentos (grupos) de la macromolécula como moléculas de referencia y se obtiene su potencial efectivo E_e calculado cuánticamente para diferentes configuraciones o mediante información proveniente de análisis experimentales ([Lipkowitz], p. 297, [McCammon], p. 38, [Gunsteren, et al, 1992], p. 946).

Otros métodos que utilizan este tipo de potenciales, pero que a diferencia de DM no consideran el tiempo, son **mecánica molecular (MM)** y los **métodos de Monte Carlo (MC)**. El método de MM es muy utilizado en simulaciones computacionales de cualquier tipo de moléculas y se basa en la búsqueda la configuración de energía potencial mínima del sistema como función de las coordenadas atómicas ([Lipkowitz], p. 298, [McCammon], p. 44). Por otra parte el método de MC es frecuentemente utilizado en estudios de mecánica estadística computacional, este método es probabilista a diferencia de la DM que es determinista ([Haile], p 13). Con los métodos de MC y DM es posible calcular propiedades termodinámicas de los sistemas mediante promedios configuracionales, en el primer método y promedios temporales en el segundo. Las propiedades termodinámicas calculadas con ambos métodos se relacionan mediante el teorema ergódico.

Inevitablemente con toda aproximación se introducen algunas limitaciones, para este tipo de aproximación clásica no es posible observar o describir adecuadamente:

- Transiciones electrónicas (absorción y emisión de fotones).
- Fenómenos de transporte electrónico (transferencia de carga).
- Formación y rompimiento de enlaces⁴.
- Efecto tunel.

⁴Esto es posible modelarlo desde un punto de vista clásico, sin embargo un método cuántico provee de información más detallada y confiable dado la naturaleza del fenómeno.

No obstante, es posible obtener aceptablemente algunos elementos importantes del comportamiento de sistemas macromoleculares que en base a un método cuántico sería prácticamente imposible y de esta manera, complementar información proveniente de métodos experimentales en donde frecuentemente hay inaccesibilidad en detalles atómicos [Gunsteren, et al, 1992].

3.3.1 Campo de fuerza para moléculas biológicas, AMBER

En la literatura encontré que campos de fuerza como los mencionados en la sección anterior son usados frecuentemente en estudios teóricos de sistemas biomoleculares ([Gunsteren, et al, 1992], [Lipkowitz], p. 295, [McCammon], p. 181), los cuales están implementados en diversos códigos para hacer simulaciones. Estos campos de fuerza han sido ajustados para representar diferentes tipos de moléculas biológicas, donde los más comúnmente usados son:

AMBER, desarrollado en la Universidad de California en San Francisco (Weiner & Kollman, 1981);

CEDAR, desarrollado en la Universidad del Norte de Carolina (Carson & Heermans, 1985 [Heermans]);

CHARMM, desarrollado en la universidad de Harvard (Brooks, 1983);

GROMOS, desarrollado en la Universidad de Groningen (van Gunsteren, 1983).

En particular AMBER es ajustado para el estudio de ácidos nucleicos y proteínas, empleando en dicho ajuste tanto métodos cuánticos como datos experimentales. Por lo anterior consideramos que es lo suficientemente completo y conveniente para usarlo en las simulaciones que se hicieron en el presente estudio. El campo de fuerzas AMBER es incorporado, con todas sus librerías, en un paquete comercial de *software* llamado *discover*, elaborado por la compañía *Biosym Tec. Inc.* [Discover]. Este paquete fue usado para el desarrollo de las dinámicas moleculares que aquí se reportan.

AMBER

El campo de fuerza AMBER fue hecho por el grupo de trabajo de Kollman y se actualiza constantemente, incluyendo modificaciones a diversos parámetros [Weiner, et al, 1984].

El potencial que genera el campo de fuerza AMBER es el siguiente:

$$\begin{aligned}
 V(r) = & \underbrace{\sum_{\text{enlaces}} K_r (R - R_{eq})^2 + \sum_{\text{angulos}} K_\theta (\theta - \theta_{eq})^2 + \sum_{\text{diedrales}} \frac{V_n}{2} [1 + \cos(n\phi - \gamma)]}_{V_{\text{ligados}}} \\
 & + \underbrace{\sum_{i < j} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^6} + \frac{q_i q_j}{\epsilon R_{ij}} \right] + \sum_{\text{puentes-H}} \left[\frac{C_{ij}}{R_{ij}^{12}} - \frac{D_{ij}}{R_{ij}^{10}} \right]}_{V_{\text{no-ligados}}} \quad (3.9)
 \end{aligned}$$

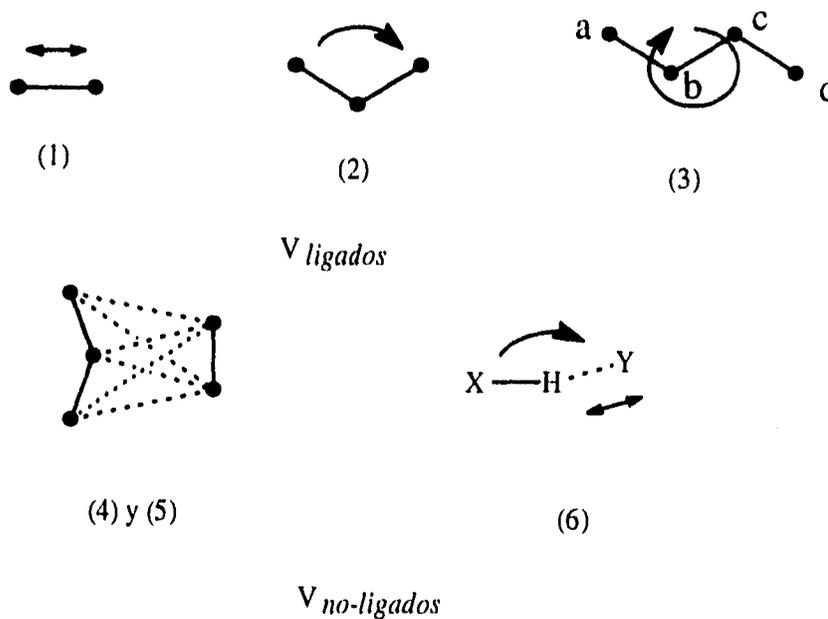


Figura 3.1: Representación esquemática de cada término del potencial de AMBER.

El término (1) representa cada enlace covalente en la molécula, el término (2) las deformaciones angulares entre dos enlaces covalentes en un átomo dado y el término (3) las torsiones (ó ángulo diedral⁵). Los restantes tres términos corresponden a la dispersión e interacciones coulombianas entre pares de átomos separados una distancia r , $V_{\text{no-ligados}}$.

⁵Ángulo formado entre los planos abc y bcd.

Los términos (1) y (2) son tratados como osciladores armónicos con **constantes de fuerza** K_r y K_θ teniendo puntos de equilibrio en R_{eq} y θ_{eq} , este término se considera que es una buena aproximación para representar enlaces covalentes en temperaturas biológicas normales, donde la amplitud de las fluctuaciones en los enlaces son muy pequeñas. El término (3) se incluye para describir y mantener apropiadamente la naturaleza estructural de las cadenas de los ácidos nucleicos y las proteínas ([McCammon], p 40). En general estos tres términos representan las variaciones energéticas en los enlaces covalentes $V_{ligados}$, es decir, la estructura primaria de los sistemas macromoleculares mencionados. Los términos (4) y (5) toman en cuenta las interacciones de van der Waals y electrostáticas. Para aumentar la naturaleza electrostática de los puentes de hidrógeno (*idem*, [McCammon] y [Pauling]), se incluye el término (6), el cual se reemplaza por el (4) en átomos de los grupos donadores y aceptores que cumplan condiciones angulares y de distancia para formar dichos puentes (fig. 3.1), en *discover* estas condiciones se fijan a 23° y 1.4 \AA respectivamente.

Revisando los artículos [Weiner, et al, 1984] y [Weiner, et al, 1986], donde se reportan las constantes de fuerza calculadas para este campo, encontré que para ajustar los parámetros de cada uno de los términos se recurre a: estudios hechos con micro-ondas, difracción de neutrones, resonancia magnética nuclear (RMN), métodos cuánticos y MM en los términos ligados y empaquetamiento cristalino en los parámetros no-ligados. Posteriormente todos los parámetros se varían y se ajustan siguiendo una metodología de tal forma que coincidan con valores calculados de energías conformacionales (en e.g. azúcar, bases, apilamiento de las bases, etc.). En los términos (1), (2) y (3), los valores de los parámetros de equilibrio R_{eq} y θ_{eq} , se toman a partir de estudios hechos con micro-ondas y difracción de neutrones, métodos cuánticos y/o MM. El γ mediante micro-ondas, RMN ó estudios MM. Los parámetros no-ligados son derivados con datos empaquetamiento cristalino. Las cargas atómicas son determinadas con análisis de población de Mulliken apéndice 6.3, de moléculas de referencia ó alternativamente en versiones más recientes de AMBER se usa el modelo de cargas atómicas parciales ajustado para reproducir el potencial electrostático, calculado cuánticamente sobre una malla de puntos alrededor de la molécula de referencia mediante cálculos electrónicos, apéndice 3.2. Los parámetros de estructura (e.g. coordenadas atómicas, ángulos entre enlaces, etc.) son frecuentemente tomados de datos de estructura cristalina, de medidas espectroscópicas o cálculos cuánticos. Actualmente los modelos permiten también la adición de solvente (iones y contraiones) para dar realismo a este tipo de simulaciones.

Parte II

Metodología y Desarrollo

Capítulo 4

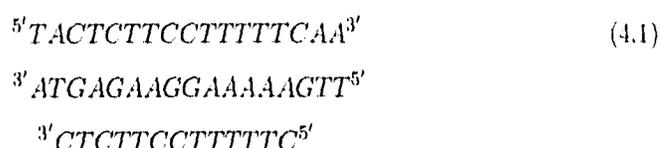
Simulación computacional

En este capítulo describo la metodología seguida para el desarrollo de la simulación, partiendo de la delimitación y construcción del sistema ADN-*triplex* propuesto para el presente estudio (cálculo de cargas atómicas parciales, cálculo de la posición de contraiones y obtención de la configuración inicial). Posteriormente presento un análisis e interpretaciones de algunos de los resultados obtenidos de la simulación.

4.1 Construcción del sistema

La secuencia de ADN-*triplex* que se muestra en (4.1), me fue sugerida para el desarrollo de la simulación. Esta se tomó del artículo de Mergny (1991), en donde se reporta un estudio de estabilidad hecho mediante el procedimiento experimental que describiré posteriormente en la sección 4.2.1. La secuencia incluye el nucleótido C⁺ (citosina protonada) para la formación de los puentes de hidrógeno tipo Hoogsteen (fig. 2.4). Debido a que no contamos con otro tipo de información estructural sobre la molécula, usé el módulo para visualización y modelado interactivo para polímeros biológicos (*Biopolymer* del paquete *Insight*), para la construcción del sistema [Insight]. Dentro de este paquete se puede controlar el programa de dinámica molecular *Discover* que como mencioné contiene en sus librerías el campo de fuerza AMBER usado para el desarrollo de la simulación. Dicho campo cuenta con parámetros y constantes de fuerza para cada átomo del ADN con sus diferentes enlaces, dadas las particularidades de esta molécula en *Insight* se utiliza una librería de residuos que incluye a los nucleótidos T, A, C y G, así como a las moléculas de agua e iones, tomadas de los parámetros calculados para el potencial de AMBER. No obstante, el nucleótido C⁺

no cuenta con cargas atómicas, por lo que fue necesario asignarlas de acuerdo con su naturaleza electrónica. La adición del proton a las citosinas del oligo (protonación), origina redistribución de carga en el nucleótido, neutralizandose la carga de este, ya que como he mencionado, los fosfatos tienen una carga formal de $-1e$.



El módulo *Biopolymer* utiliza parámetros cristalográficos obtenidos por difracción de fibra [Arnott, et al, 1974-76], para las uniones triples T:A:T como se ha mencionado. Para las uniones C:G:C⁺, debido a que sus librerías no cuentan con información cristalográfica de ellos, las construí con dicho módulo remplazando las bases T:A:T por C:G:C en los sitios que indica la secuencia, fijando el pH requerido mediante la protonación de las citosinas del oligo con la hibridización sp²⁺ del nitrógeno N3 (fig. 4.1), formándose de esta manera los puentes de hidrógeno tipo *Hoogsteen*. Las constantes de fuerza para el nucleótido C⁺, las tomé de las librerías de AMBER usando para ello la simbología descrita en [Weiner, et al, 1984], [Weiner, et al, 1986] y [Discover], esta simbología se utiliza para representar los diferentes tipos de átomos dadas las características químicas en las que se encuentran¹, (tabla 4.1). Debido a que las librerías de AMBER no cuentan con cargas atómicas para el nucleótido C⁺, estas las calculé mediante el procedimiento descrito en la sección 4.1.1.

Atomos	Orden de enlace	Potencial de Amber
N3 C2	1 (simple)	N (nitrogeno sp ² en un grupo amide)
H3 N3	1 (simple)	H (hidrogeno amide o imino)
C4 N3	2 (doble)	CM (carhono sp ² en pirimidina, un substituyente)

Tabla 4.1: Potenciales de AMBER asignados a la Citosina protonada.

Posteriormente, saturé los enlaces libres de los extremos de las cadenas mediante un procedimiento de *Biopolymer*, el cual reemplaza el fosfodiéster por un radical OH en 5' y agrega el mismo radical en 3'. Finalmente incluí contraiones y agua como solvente llenando una celda unitaria de $50 \times 50 \times 80 \text{ \AA}^3$, siguiendo los pasos descritos en las secciones 4.1.2 y 4.1.3 respectivamente. Para ajustar la estructura en base a las interacciones de todos los

¹Propiedades estereoquímicas.

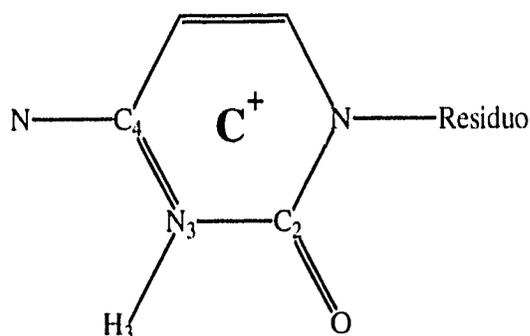


Figura 4.1: Citosina protonada, H3 es el proton que se agrega a la molécula.

elementos del sistema se siguió el procedimiento descrito en la sección 4.1.3 obteniendo así una configuración de mínima energía (fig. 4.3 a).

4.1.1 Cálculo de cargas atómicas para C⁺

Inicialmente calculé las cargas atómicas con el análisis de población de Mulliken, apéndice 6.3, debido a que es la forma en que se hizo en la versión original de AMBER [Weiner, et al, 1984]. Sin embargo, las librerías de AMBER que usa *Discover* corresponden a versiones actualizadas a 6-mayo-94, en donde las cargas atómicas son ajustadas de tal forma que reproducen el potencial electrostático, apéndice 6.4 calculando dicho potencial con métodos *ab-initio* como mencioné en el capítulo 3. Considerando que el valor de las cargas varía de un método a otro [Singh, et al, 1984], para mantener congruencia con la distribución de cargas que usamos en la simulación fue necesario recalcularlas usando el potencial electrostático. En la tabla 4.2 se muestran los resultados que se obtuvieron de las cargas atómicas para la citosina protonada con ambos métodos, también se muestran las cargas parciales de AMBER para la citosina sin protonar calculada en [Weiner, et al, 1986] (versión que usa *Discover*) y las obtenidas con métodos semiempíricos [Laughton, et al, 1992] mediante el procedimiento propuesto por Orozco y Luque [Orozco, et al, 1990] (en este último no se reportan cargas para el grupo fosfato), con objeto de compararlas.

Comparando los valores de las cargas atómicas obtenidas, tabla 4.2, entre los diferentes métodos reportados son notorias las diferencias entre los átomos, como se esperaba. Por lo que al utilizar el mismo método de AMBER para calcular las cargas del nucleótido

C^+ su valor se mantendrá congruente, puesto que en el cálculo se considera de la misma forma el ambiente químico que rodea cada átomo de C^+ que en los nucleótidos A, T, G ó C, consecuentemente el campo electrostático generado representará homogéneamente las distribuciones de carga que produce la presencia de los nucleótidos C^+ a lo largo de la molécula. No obstante, aún cuando las cargas atómicas son de relevancia fundamental para este tipo de simulaciones es por sí mismo un completo tema de estudio y profundizar más en este aspecto estaría fuera de los objetivos de lo acordado en el plan de trabajo de la presente tesis, dejándolo por el momento para discusiones ó estudios posteriores, ya que considero que dada la naturaleza electrostática, tanto de puentes de hidrógeno, de la ubicación de los iones presentes en el medio como de la repulsión de las cadenas es un punto muy importante que debe retomarse para tener mejores representaciones.

4.1.2 Posición de contraiones

Considerando que la posición de un contraion está determinada principalmente por su carácter electrostático, para determinar su localización inicial usé las cargas atómicas de la molécula de ADN, como una distribución de cargas puntuales q_1, q_2, \dots, q_n localizadas en $\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_n$, para calcular el potencial electrostático generado por todo el ADN, mediante la expresión:

$$Ve(\mathbf{r}) = \frac{1}{4\pi\epsilon_0} \sum_{i=1}^N \frac{q_i}{\|\mathbf{r} - \mathbf{r}_i\|}.$$

De esta forma en los mínimos del potencial generado por dicha distribución coloco los contraiones con objeto de obtener una configuración inicial "razonable" a partir de la cual se pueda iniciar un proceso de minimización de energía que incluya todos los términos del potencial. Para esto procedí de la siguiente manera:

Tomando puntos sobre la superficie cilíndrica dada por.

$$\begin{aligned} x &= \rho \cos \theta, \\ y &= \rho \sin \theta, \\ z &= z. \end{aligned} \tag{4.2}$$

Con $\rho = cte$ tal que contenga totalmente a la molécula encerrada en una superficie reticular cilíndrica, podemos evaluar el valor del potencial sobre dicha retícula y buscar la posición de los mínimos de potencial electrostático, tomando la retícula lo más densa posible

	A	cap 3'	cap 5'	B	C	cap 3'	cap 5'	D	cap 3'	cap 5'
P	1.385	1.385			1.513	1.513		1.385	1.385	
O1P	-0.847	-0.847			-0.814	-0.814		-0.847	-0.847	
O2P	-0.847	-0.847			-0.831	-0.831		-0.847	-0.847	
O5'	-0.509	-0.509	-0.404	-0.659	-0.747	-0.747	-0.747	-0.632	-0.632	-0.558
H5'			0.279				0.417			0.346
C5'	0.118	0.118	0.118	0.213	0.033	0.033	0.033	0.299	0.299	0.299
H5'1	0.021	0.021	0.021	0.044	0.204	0.204	0.204	0.041	0.041	0.041
H5'2	0.021	0.021	0.021	0.044	0.210	0.210	0.210	0.041	0.041	0.041
C4'	0.036	0.036	0.036	0.095	0.082	0.082	0.082	0.146	0.146	0.146
H4'	0.056	0.056	0.056	0.082	0.231	0.231	0.231	0.068	0.068	0.068
C3'	0.233	0.233	0.233	0.130	0.144	0.144	0.144	0.581	0.581	0.581
H3'	0.025	0.025	0.025	0.143	0.158	0.158	0.158	-0.091	-0.091	-0.091
O4'	-0.368	-0.368	-0.368	-0.391	-0.676	-0.676	-0.676	-0.558	-0.558	-0.558
C1'	0.376	0.376	0.376	0.255	0.385	0.385	0.385	0.530	0.530	0.530
H1'	0.009	0.009	0.009	0.188	0.273	0.273	0.273	0.073	0.073	0.073
C2'	-0.307	-0.307	-0.307	-0.308	-0.403	-0.403	-0.403	-0.360	-0.360	-0.360
H2'1	0.081	0.081	0.081	0.155	0.249	0.249	0.249	0.140	0.140	0.140
H2'2	0.081	0.081	0.081	0.155	0.204	0.204	0.204	0.140	0.140	0.140
O3'	-0.509	-0.508	-0.509	-0.641	-0.749	-0.749	-0.749	-0.756	-0.756	-0.756
HO3'		0.306				0.454		0.455		
N1	-0.187	-0.187	-0.187	-0.711	-0.800	-0.800	-0.800	-0.710	-0.710	-0.710
C2	0.859	0.859	0.859	1.065	1.100	1.100	1.100	1.024	1.024	1.024
O2	-0.508	-0.508	-0.508	-0.494	-0.592	-0.592	-0.592	-0.601	-0.601	-0.601
N3	-0.860	-0.860	-0.860	-1.034	-0.914	-0.914	-0.914	-0.803	-0.803	-0.803
H3				0.507	0.473	0.473	0.473	0.450	0.450	0.450
C4	0.935	0.935	0.935	1.309	0.833	0.833	0.833	1.171	1.171	1.171
N4	-0.834	-0.834	-0.834	-1.257	-0.922	-0.922	-0.922	-1.167	-1.167	-1.167
H41	0.329	0.329	0.329	0.577	0.443	0.443	0.443	0.527	0.527	0.527
H42	0.351	0.351	0.351	0.618	0.463	0.463	0.463	0.538	0.538	0.538
C5	-0.576	-0.576	-0.576	-1.181	-0.398	-0.398	-0.398	-0.753	-0.753	-0.753
H5	0.153	0.153	0.153	0.339	0.296	0.296	0.296	0.264	0.264	0.264
C6	0.185	0.185	0.185	0.924	0.217	0.217	0.217	0.472	0.472	0.472
H6	0.098	0.098	0.098	0.192	0.331	0.331	0.331	0.152	0.152	0.152
total	-1.000	-0.693	-0.307	0.359	-0.002	0.452	0.547	-0.0827	0.3719	0.646
3' + 5'			-1.000				0.999			1.018

Tabla 4.2: Cargas atómicas parciales calculadas **A** en AMBER para la citosina; **B** en el artículo Laughton, **C** mediante el Análisis de Población de Mulliken y **D** mediante el potencial electrostático ChelpG, estas últimas para la citosina protonada. Los cap son usados para los extremos de las cadenas.

para que los resultados sean más exactos, en este caso tomé $dz=0.5 \text{ \AA}$ y $ds=\rho d\theta = 0.5$, dado que el radio iónico del sodio es de ~ 1.5 y sólo se busca una posición aproximada para distribuir los contraiones.

Para simplificar la tarea escribí un programa en Fortran que consta de dos fases:

- En la primera se calcula una posición aproximada de cada contraion debida a todas las cargas, incluyendo los contraiones que ya se han localizado en cada iteración.
- En la segunda se utiliza un método autoconsistente, en el cual cada contraion se quita y se vuelve a reubicar, buscando el potencial electrostático mínimo en el campo producido por el ADN y los contraiones restantes. Esto se hace para cada uno de los contraiones hasta que la distancia entre la posición anterior y la que está calculando, no exceda 1 \AA de diferencia. Refinándose la posición de los contraiones en cada iteración y evitando que el contraion se ubique en una posición ficticia² de potencial electrostático mínimo.

Estos cálculos se hicieron para superficies cilíndricas con $13 \text{ \AA} < \rho < 17 \text{ \AA}$ observándose que a 15 \AA los contraiones se ordenan a la altura de cada fosfato a lo largo de las cadenas de la *duplex* y su número coincide aproximadamente (± 1) con la carga de cada una de ellas. Para el oligo los contraiones se ubican en diferentes posiciones, lo cual es de esperarse, debido a que esta cadena contiene las protonaciones de C^+ (cargas positivas) en distintos sitios. Esto concuerda cualitativamente con lo reportado en el artículo [Laughton, et al, 1992].

El procedimiento para determinar el potencial electrostático mínimo descrito anteriormente fue verificado usando otros métodos (e.g., *Steepest Descents*, Gradientes y criterio de Hessiano) y coinciden dentro del error de la redícula seleccionada. Para hacer esta comparación se escribieron programas en Fortran basados en los algoritmos antes mencionados [Hostetter]. Sin embargo el método utilizado consume un menor tiempo de cómputo, lo cual es útil para el estudio de secuencias mayores o para procedimientos rutinarios.

4.1.3 Configuración inicial

La forma en que añadí los contraiones sólo toma en cuenta el término electrostático del potencial. Desafortunadamente la forma en que *Insight* añade el solvente, aún cuando

²Ficticia, en el sentido de que pudo haber sido una posición mínima en un principio, sin embargo considerando la presencia de todos los contraiones ya no lo sea más, cambiando su posición.

toma en cuenta la densidad del agua, no considera la presencia de la molécula *triplex* y contraiones bajo ninguna interacción (únicamente empalme de átomos). Por lo que consideré conveniente distribuir el solvente, los contraiones y finalmente la molécula *triplex*, permitiendo que cada uno se ajuste bajo la influencia del otro de la siguiente manera:

- Eliminé 43 moléculas de agua cercanas a la molécula de ADN ($<2.0 \text{ \AA}$). puesto que estas pueden quedar atrapadas en su interior y deformar la estructura durante la simulación [Discover].
- Minimización de energía de las moléculas de agua manteniendo fija la estructura del ADN y los contraiones. Usando para ello los algoritmos *Steepest decents* y Gradientes Conjugados por 150 y 600 iteraciones, respectivamente.
- Minimización de energía en la que permitió movilidad a los iones, para que estos se ajusten junto con el agua a la estructura del ADN, usando los algoritmos *Steepest decents* y Gradientes Conjugados por 150 y 400 iteraciones respectivamente.
- Una pequeña dinámica del solvente a baja temperatura, con el único objeto de redistribuir las moléculas de agua e iones agregando un poco de energía cinética en ambos para "brincar" pequeñas barreras de potencial manteniendo fijo el ADN. Con 500fs de inicialización y 5ps de dinámica a una temperatura de 100K, con un ensamble NPT para ajustar el tamaño de la celda.
- Minimización global del sistema, 100 iteraciones de *Steepest decents* y 900 de Gradientes Conjugados. Obteniendo una desviación cuadrática media de la derivada de 0.1 kcal/mol\AA , sin fijar ninguna parte del sistema (fig. 4.3 a)).

El procedimiento anterior se determinó a partir de simulaciones de prueba y de resultados de otras simulaciones reportados en los artículos, [Laughton, et al, 1992] y [Swaminathan, et al, 1991], del libro [McCammon] y [Discover].

4.1.4 Simulación con dinámica molecular

Con objeto de estudiar la estabilidad de las estructuras se desarrollaron simulaciones a dos diferentes temperaturas y con diferentes ensambles, apéndice 6.5. Las temperaturas se escogieron a partir de las observaciones experimentales reportadas por Mergny

[Mergny, et al, 1991] para esta secuencia, en donde a 293 K la estructura de *triplex* es estable y a 323 K hay desnaturalización de la estructura *triplex*, (fig. 4.5). Haciendo una comparación visual de las simulaciones NVT a 293 K y NPT a 293 K, ambas 2.5 ps de inicialización y 36.2 ps de dinámica, se observa que las moléculas de agua en el ensamble NVT se alejan demasiado de el ADN dejando un vacío de más de 4 Å. Debido a lo anterior, me atrevo a decir que la presión es un factor importante para simular correctamente las condiciones bajo las cuales se encuentra el ADN ya que algunos grupos del ADN son hidrofóbicos [Saenger] y con suficiente espacio disponible en NVT, las moléculas de agua se alejan de él. Por lo cual supongo que las simulaciones con un ensamble NPT son más convenientes. Para iniciar la simulación fue necesario fijar diversos parámetros, puesto que es la primera simulación que hago y no tengo experiencia al respecto, recurrí a algunas sugerencias, datos reportados en artículos y simulaciones previas. Para simular el ensamble NPT es necesario incluir condiciones periódicas y una distancia de corte (*cutoff*) importante para poner un límite en los cálculos (fig. 4.3 b)). Las condiciones periódicas usadas para estas simulaciones usan la convención de imagen mínima, es decir, solo se considera una vez las interacciones de cada átomo con el resto de los átomos que estén dentro de la distancia de corte, ya sea con un átomo real o su imagen. Para evitar cortes abruptos en la energía se usa una función interruptor (*switch*), la cual vale uno hasta cierta distancia (*cutdis-swtdis*), después es llevada a cero suavemente hasta un valor fijo antes de la distancia de corte (*cutdis*), con un polinomio de quinto grado (fig. 4.2). Esta distancia se recomienda que se fije midiendo las distancias entre los grupos cargados [Discover], por lo que para nuestra molécula los valores que fijé son: *cutoff*=13 Å, *cutdis*=12 Å y *swtdis*=3.5 Å.

De acuerdo con lo anterior se llevaron al cabo las siguientes simulaciones:

1. NPT, 293 K con 2.5ps de inicialización y 36.2ps de dinámica.
2. NPT, 323 K con 5.5ps de inicialización y 9.2ps de dinámica.

Los tiempos de inicialización me fueron sugeridos tomando en cuenta que: "a mayor temperatura el sistema tarda más en estabilizarse".

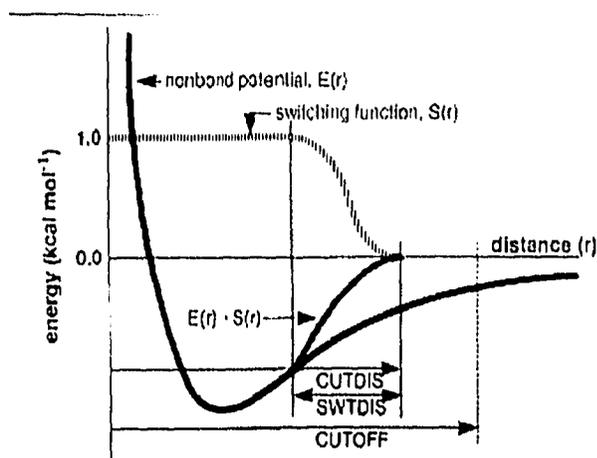
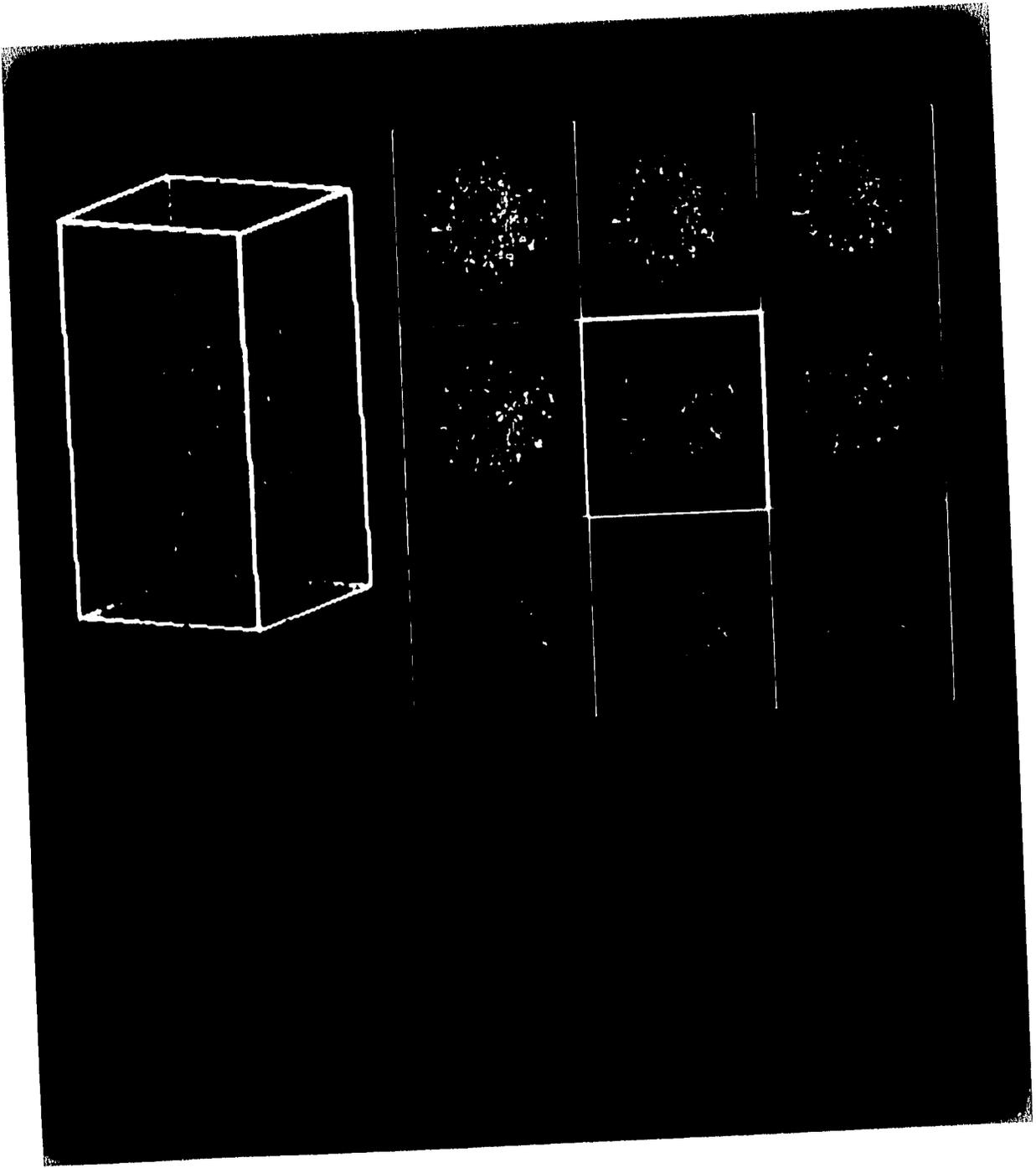


Figura 4.2: Gráfica que esquematiza el efecto de la función interruptor (switch) en la energía. Tomado de [discover].

4.2 Análisis e interpretación de resultados

En la presente sección hago un pequeño análisis de cuatro puntos que considero importantes sobre los resultados de la simulación, presentando también la interpretación que tengo de ellos. Debo mencionar que esto ha sido el fruto de la revisión del material bibliográfico y de múltiples discusiones que se han hecho al respecto del tipo de información que se pueda extraer de la simulación. Tanto el análisis como la interpretación tienen el fin de estudiar factores relacionados con la estabilidad de la estructura secundaria del ADN-*triplex*, correspondiente con lo esbozado en el plan de trabajo de la tesis.

Antes de iniciar el análisis es conveniente observar la gráfica de energía potencial del sistema a lo largo de la simulación (fig. 4.4), ya que para asegurar el equilibrio termodinámico del sistema es necesaria su estabilidad energética. La gráfica fue calculada con las posiciones y velocidades atómicas de las simulaciones a 293 K y 323 K para cada tiempo usando para ello el paquete *Insight*. La energía potencial del sistema en ambas temperaturas no ha llegado a un valor constante, sin embargo tiende a un valor muy cercano a la asíntota de la curva ajustada por mínimos cuadrados. Este punto lo menciono, por que pienso que es conveniente aumentar el tiempo de equilibrio (inicialización) en simulaciones posteriores. No obstante, consideramos aceptables estos valores ya que existen simulaciones más largas



(ns), e. g. [Swaminathan, et al, 1991], en donde aún a estos tiempos se observa un pequeño decaimiento exponencial hacia la constante.

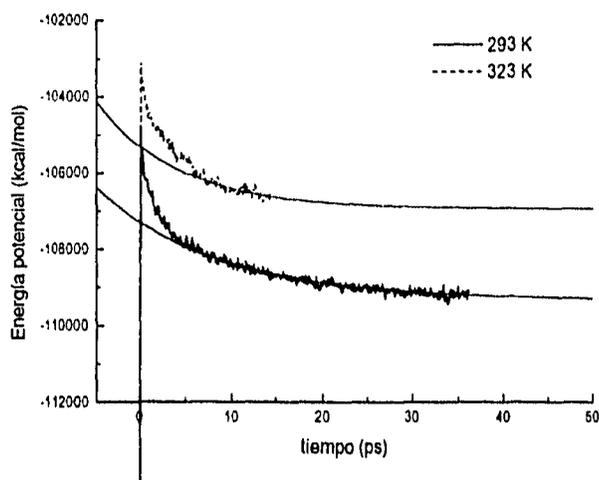


Figura 4.4: Energía potencial del sistema a las temperaturas indicadas.

4.2.1 Parámetros intermoleculares

Con el propósito de comparar la estabilidad de la estructura secundaria del ADN-triplex a diferentes temperaturas, calculé la energía de interacción intermolecular entre las cadenas, el número y distancias de separación de los puentes de hidrógeno formados entre las bases. Para ello, escribí un programa macro con el lenguaje BCL (*Biosym command language*) de *Insight* y usé un programa en Fortran que calcula la energía intermolecular del término 10-12 de la ecuación (3.9) usando las constantes de fuerza de las librerías de AMBER. Los promedios temporales de estos parámetros obtenidos después de la inicialización se muestran la tabla 4.3. A partir de estos parámetros, se observa que a 293 K el número de puentes formados entre las cadenas 1-3 es mayor que a 323 K, lo mismo que la energía de interacción, en donde la diferencia es notable comparada con cualquier valor obtenido, por lo que es posible que este hecho se relacione mucho con la desnaturalización del ADN-triplex (fig. 4.5). Por otra parte, la energía total intermolecular entre las cadenas 1-3 es mayor que entre 1-2 a 293 K, ocurriendo lo contrario a 323 K. Así también a 293 K el

número de puentes de hidrógeno formados entre las cadenas 1-2 es menor que a 323 K. Estas dos observaciones pueden estar relacionados con la desestabilización local que la presencia del oligo haga sobre la estructura *duplex*. Las distancias promedio de los puentes de hidrógeno no nos proporcionan mucha información al respecto, debido a que son promedios sobre los puentes de hidrógeno formados bajo el criterio de distancia y ángulo mencionado en modelado molecular.

Propiedades	Moléculas 1-2	Moléculas 1-3	Moléculas 1-2	Moléculas 1-3
	293 K	293 K	323 K	323 K
Coulomb (Kcal/mol)	139.29	79.01	130.37	77.34
E Puentes de H. (Kcal/mol)	-10.71	-8.33	-11.51	-6.81
Van der Wals (Kcal/mol)	-25.19	61.59	-22.53	-68.86
E. Total. (Kcal/mol)	103.39	132.27	96.33	1.67
Num. de Puentes de H.	45	28	49	24
Dist. Promedio (Å)	2.08	2.06	2.10	2.04

Tabla 4.3: Promedios temporales de las energías intermoleculares y parámetros correspondientes a los puentes de hidrógeno de las dinámicas a 293 K y 323 K

En la simulación a 323 K no observamos la separación del oligo (desnaturalización), debido a que este proceso experimentalmente toma del orden de segundos [McCammon], aun cuando se caliente subitamente la solución. No obstante, con una dinámica de corto tiempo, es posible describir algunos elementos del proceso de desnaturalización, lo cual puede ser aprovechable para diferenciar entre oligos más o menos estables de forma analoga a la experimental.

4.2.2 Estructura secundaria del ADN-*triplex*

Para observar la flexibilidad de las cadenas del ADN calculé los ángulos de torsión definidos a lo largo de las cadenas y en el azúcar (tabla 4.4 y fig. 4.6). Adicionalmente con estos valores se puede determinar la forma de la estructura secundaria que adopta, para clasificar al ADN como he mencionado en la sección 2.2 (polimorfismo estructural).

Las diferentes conformaciones que puede tener el azúcar para geometrías ideales del anillo, pueden describirse de forma analítica con el **ciclo de pseudorotación** [Altona, et al, 1974-94]. Esto es debido a la forma pentagonal que forman sus enlaces cuyos ángulos de torsión se encuentran altamente correlacionados, de tal forma que es posible comprobar que:

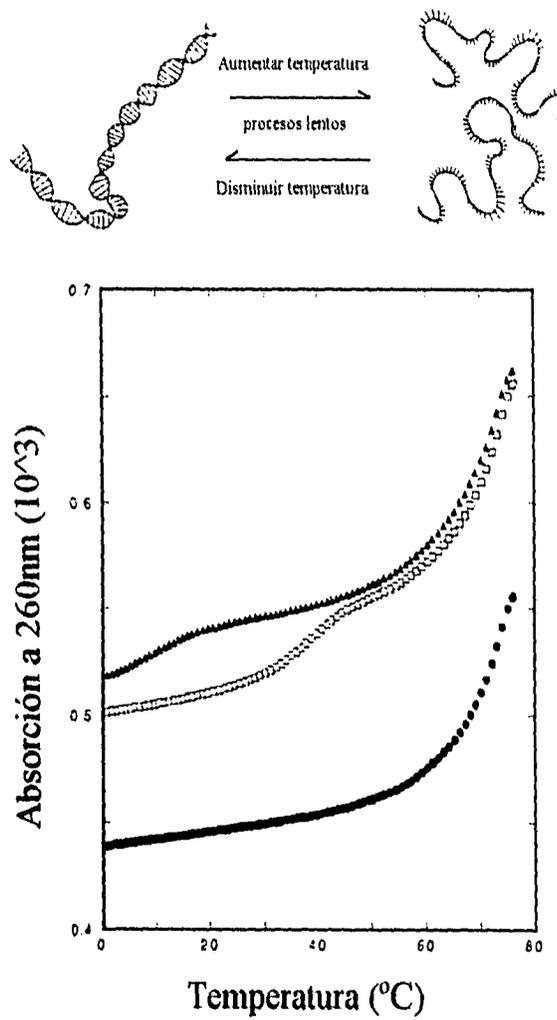


Figura 4.5: Absorción óptica a 260nm como función de la temperatura de ADN-duplex ●, ADN-triplex □ y triplex con mismatch △. Con la absorción óptica es posible determinar la desnaturalización debido al hiper Cromismo observado a ciertas temperaturas, en □ también se observa la transición de triplex a duplex. Tomado de [Mergny].

ángulo	átomos involucrados
α	$(n-1)O_{3'} - P - O_{5'} - C_{5'}$
β	$P - O_{5'} - C_{5'} - C_{4'}$
γ	$O_{5'} - C_{5'} - C_{4'} - C_{3'}$
δ	$C_{5'} - C_{4'} - C_{3'} - O_{3'}$
ϵ	$C_{4'} - C_{3'} - O_{3'} - P$
ζ	$C_{3'} - O_{3'} - P - O_{5'(n+1)}$
χ	$O_{4'} - C_{1'} - N_1 - C_2$ (pirimidinas) $O_{4'} - C_{1'} - N_9 - C_4$ (purinas)
ν_0	$C_{4'} - O_{4'} - C_{1'} - C_{2'}$
ν_1	$O_{4'} - C_{1'} - C_{2'} - C_{3'}$
ν_2	$C_{1'} - C_{2'} - C_{3'} - C_{4'}$
ν_3	$C_{2'} - C_{3'} - C_{4'} - O_{4'}$
ν_4	$C_{3'} - C_{4'} - O_{4'} - C_{1'}$

Tabla 4.4: Definición de ángulos de torsión a lo largo de la estructura primaria del ADN.

$$\nu_0 + \nu_1 + \nu_2 + \nu_3 + \nu_4 = 0,$$

$$\nu_j = \nu_{\max} \cdot \cos(P + j \cdot \varphi),$$

$$\nu_{\max} = \nu_0 / \cos P.$$

La definición original de ángulo fase es dada en términos de las torsiones del anillo de azúcar, donde ν_{\max} es la amplitud máxima, ν_j es el j-ésimo ángulo de torsión ($j=0,4$), P es el ángulo fase y $\varphi = 144^\circ$, [McCammon].

P se define como:

$$\tan P = \frac{(\nu_4 + \nu_1) - (\nu_3 + \nu_0)}{2 \cdot \nu_2 \cdot (\sin 36^\circ + \sin 72^\circ)}. \quad (4.3)$$

El ángulo fase $P=0^\circ$ es tomado como el máximo ángulo de torsión positivo de ν_2 [Saenger].

En las figuras 4.7, 4.8 y 4.9, se muestran las gráficas de los ángulos de torsión y pseudorotación tomados de la simulación a 293 K para cadenas 1, 2 y 3 respectivamente. En estas gráficas la variación radial representa el tiempo y la angular la torsión ó pseudorotación correspondientes a cada uno de los nucleótidos de las cadenas. Para hacer estas gráficas escribí un programa macro en BCL que mide los ángulos de torsión usando *decipher* (programa manejado por *Insight*) junto con otro programa en C que calcula el ángulo P y hace las gráficas usando las librerías *XGL* del sistema operativo de UNIX.

En cuanto a la flexibilidad de las cadenas podemos observar que existe poca variación en la amplitud y que fluctua alrededor de valores medios en todas las cadenas, sin embargo

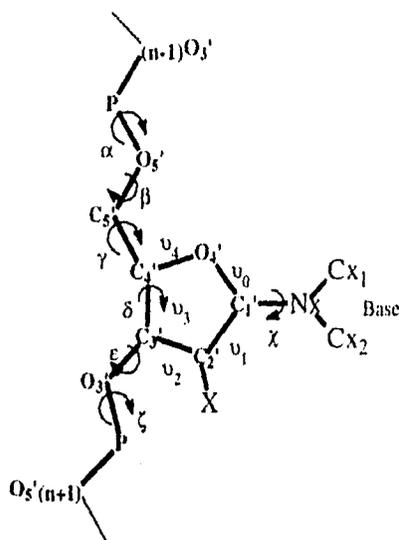


Figura 4.6: Esquema que representa los ángulos de torsión.

en la cadena 3 (fig. 4.9) se observan oscilaciones muy amplias en el ángulo P y en menor grado en el ángulo α de algunos nucleótidos. Comparando nuestros resultados con los valores descritos para un A-ADN y un B-ADN [Olson], que se muestran en la parte inferior de las figuras, las oscilaciones de los nucleótidos T2 y C8 en el ángulo P ocurren entre las conformaciones del azúcar C_3' -endo \rightleftharpoons O_4' -endo \rightleftharpoons C_2' -endo (fig. 4.10), siendo la primera y la última común para estructuras A y B-ADN, y la segunda es un punto de transición inestable energéticamente entre las otras dos configuraciones (fig. 4.11) [Saenger].

Al respecto de la forma de la estructura secundaria que adopta el ADN-*triplex*, durante la simulación la mayoría de los ángulos de torsión y pseudorotación de los nucleótidos se mantienen en un A-ADN, estructura que se construyó inicialmente. Sin embargo, es notorio que en la cadena 2 (fig. 4.8) se dan transiciones, aparentemente correlacionadas en tiempo, de los ángulos de torsión y pseudorotación a un B-ADN únicamente en los nucleótidos T6 y T10 que tienen una vecindad con C5 y C9 respectivamente, por lo que supongo que este cambio de conformación depende fuertemente de la secuencia y ya que sucedan prácticamente al mismo tiempo, es posible que una transición motive a la otra [McCammon]. Con estudios de resonancia magnética nuclear (RMN)³, se han determinado

³Esto se infiere, a partir de la determinación de la distancia relativa interprotónica entre las bases usando

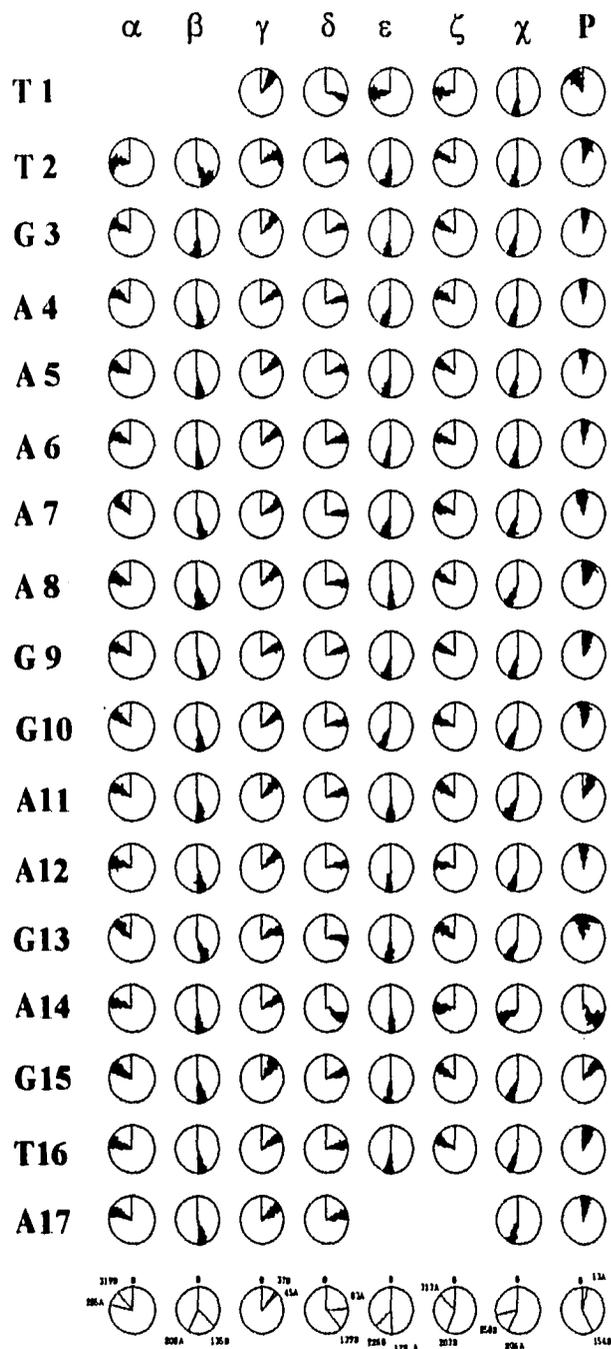


Figura 4.7: Ángulos de torsión y pseudorotación obtenidos para la cadena 1 a 293 K. En la parte inferior se muestran los mismos ángulos característicos de las estructuras secundarias A y B-ADN [Olson].

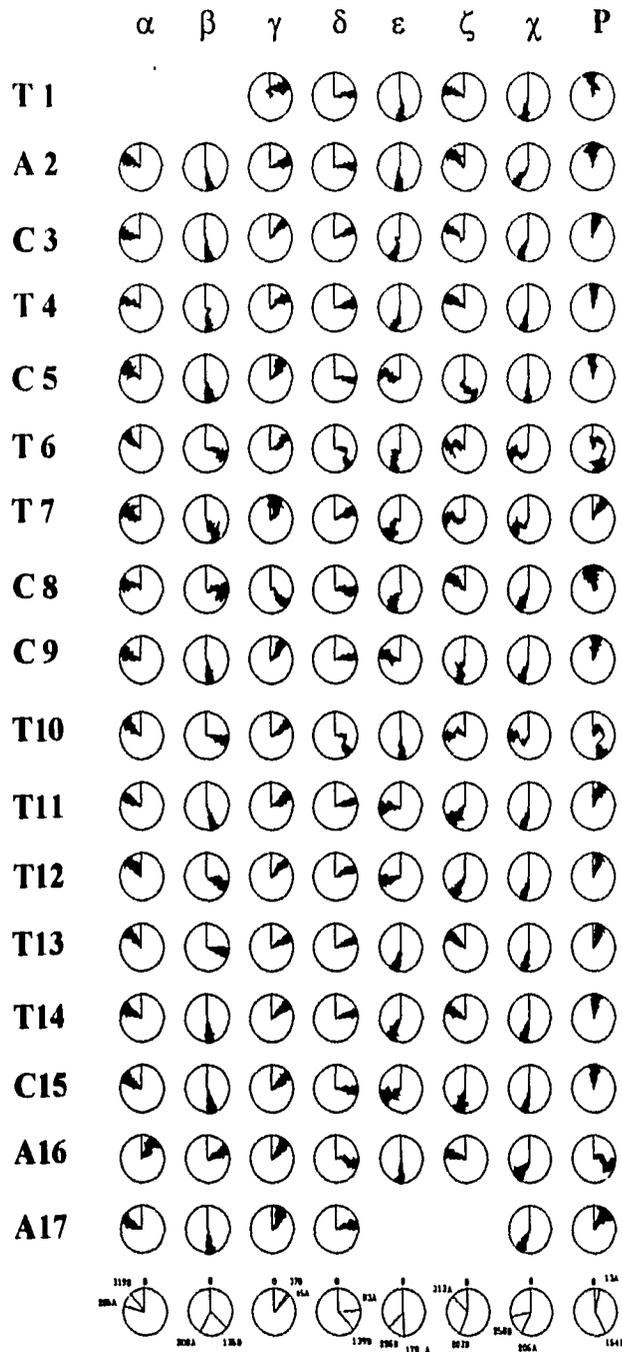


Figura 4.8: Ángulos de torsión y pseudorotación obtenidos para la cadena 2 a 293 K. En la parte inferior se muestran los mismos ángulos característicos de las estructuras secundarias A y B-ADN [Olson].

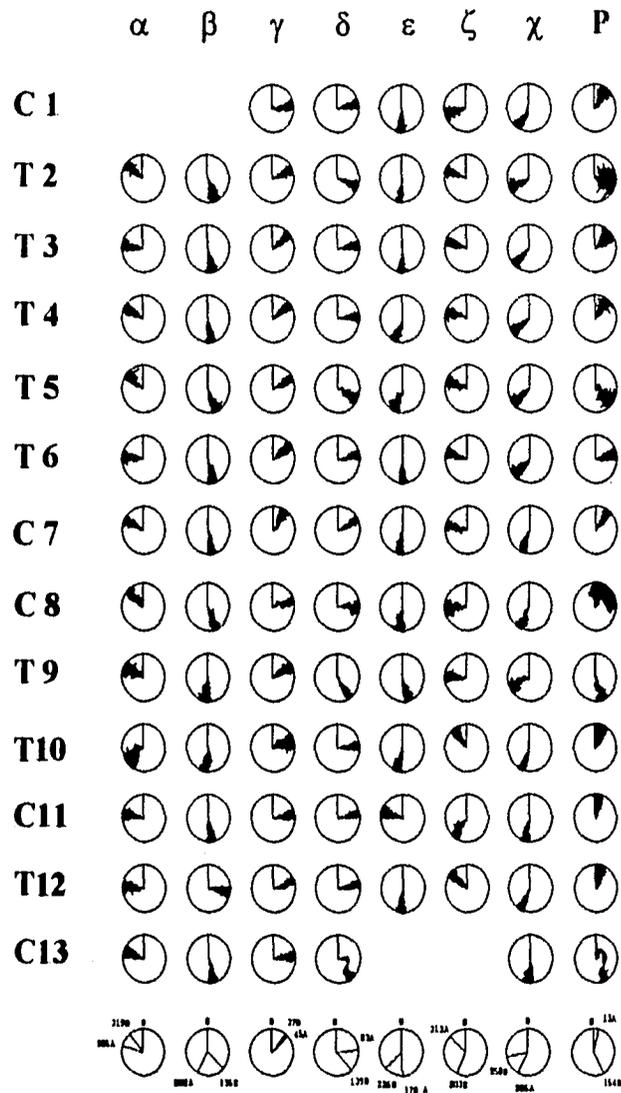


Figura 4.9: Angulos de torsión y pseudorotación obtenidos para la cadena 3 a 293 K. En la parte inferior se muestran los mismos ángulos característicos de las estructuras secundarias A y B-ADN [Olson].

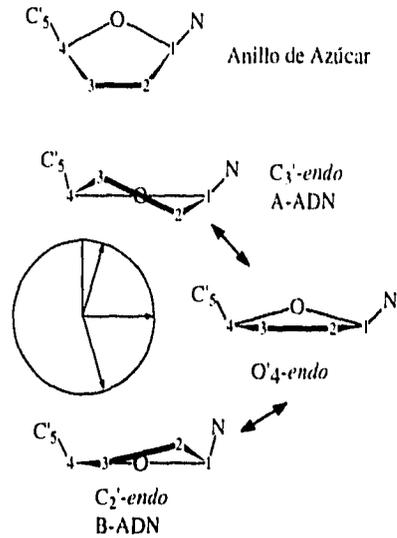


Figura 4.10: Esquema que representa la transición entre distintas configuraciones del azúcar, descritas mediante el ángulo de pseudorotación [Saenger].

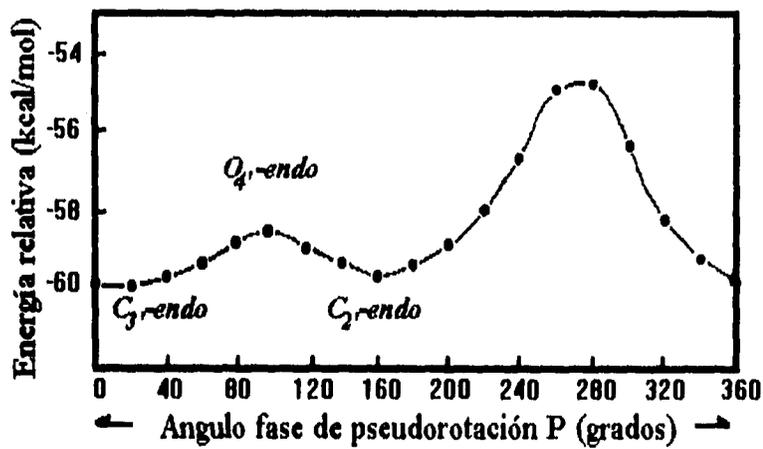


Figura 4.11: Energía vs. ángulo P, en ella se observan dos mínimos de energía que coinciden con las configuraciones obtenidas en nuestra simulación [Saenger].

conformaciones del azúcar correspondientes a un B-ADN (C_2' -endo) en estructuras *triplex* que contienen tripletes C:G:C⁺ [Laughton, et al, 1992]. Estas variaciones en la conformación del azúcar es un punto de actual controversia en investigaciones experimentales; por una parte debido a que en el modelo de difracción de fibras todas las azúcares tienen una estructura tipo A-ADN (la conformación C_3' -endo) [Arnott, et al, 1974-76] y por otra a que en estudios de RMN se encuentran algunas correspondientes al tipo B-ADN (C_2' -endo), lo cual es aceptable ya que ambos puntos son estables energéticamente como mencioné anteriormente (fig. 4.11).

Para 323 K se puede observar esencialmente lo mismo que en las gráficas descritas anteriormente, incluso las transiciones que dependen de la secuencia. No obstante, dado a que en esta temperatura experimentalmente se sabe que el oligo se "despega" de las cadenas, yo esperaré observar fuertes oscilaciones o valores fuera de los característicos de los ángulos de torsión, esto confirma que para observar completamente un proceso de desnaturalización es necesario hacer simulaciones de mayor tiempo (figs. 4.12, 4.13 y 4.14).

4.2.3 Estructura electrónica molecular

Los cálculos de estructura electrónica molecular a nivel cuántico, los utilicé para estudiar las características del enlace formado por los puentes de hidrógeno correspondientes a la unión entre las bases tipo *Hoogsteen*, en distintos segmentos de la molécula. Para este estudio eliminé a las bases los grupos fosfato, saturando con hidrógenos todos los enlaces abiertos, por lo que en el cálculo no son considerados los grupos cargados de las moléculas. Los cálculos se realizaron a nivel semiempírico empleando los parámetros AM1 implementados en el programa UNICHEMTM, se eligió este método debido a la cantidad de átomos utilizados (~90 átomos) y a que los parámetros han sido ajustados para reproducir correctamente las interacciones atómicas de muchos sistemas biológicos, incluyendo la formación de los puentes de hidrógeno intermoleculares ([Levine], p582).

En la tabla 4.5 se muestran las energías de las bases de *triplex* mencionadas, cuyas coordenadas son las obtenidas de la configuración inicial descrita en la sección 4.1.3.

Observamos que las energías totales para el T:A:T en los sitios a) y b) (los cuales son vecinos) son muy similares, sin embargo la energía en la posición d) de la misma unión es mayor que las anteriores, pero a diferencia este tiene un vecino C:G:C⁺. Por lo que el efecto NOE (Nuclear Overhauser Effect) entre sus protones aromáticos.

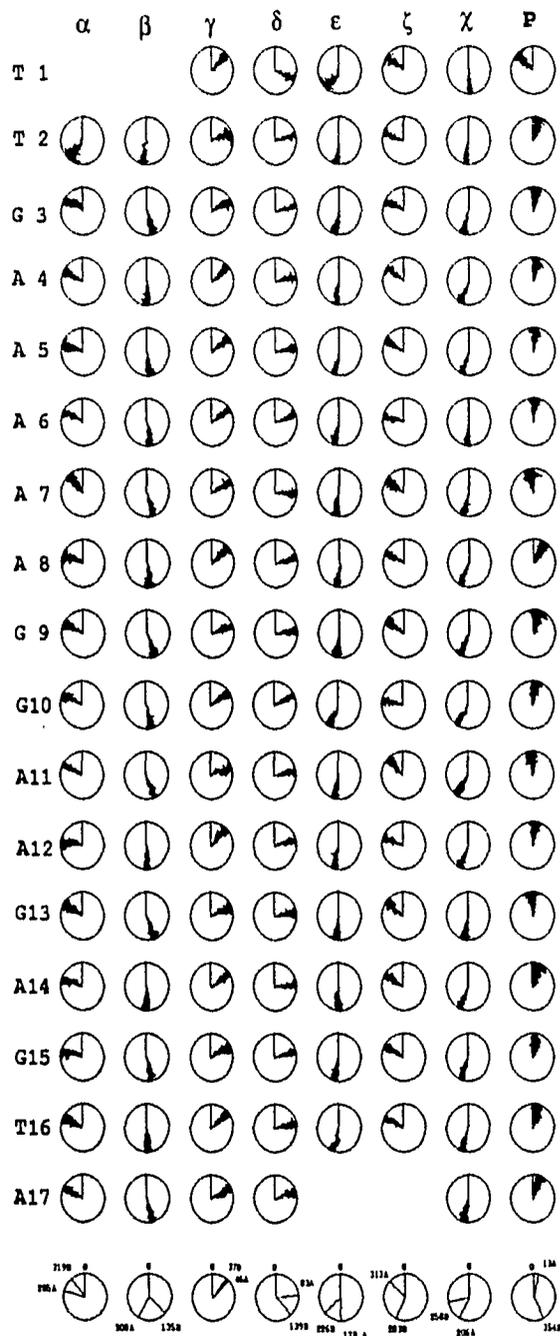


Figura 4.12: Ángulos de torsión y pseudorotación obtenidos para la cadena 1 a 323 K. En la parte inferior se muestran los mismos ángulos característicos de las estructuras secundarias A y B-ADN [Olson].

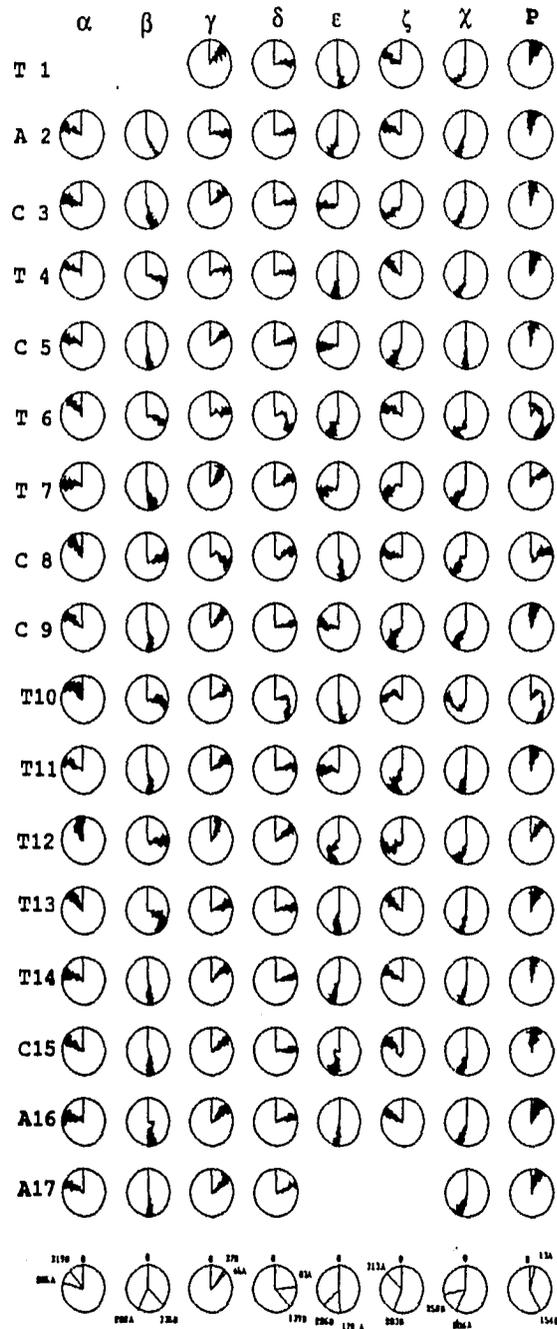


Figura 4.13: Ángulos de torsión y pseudorotación obtenidos para la cadena 2 a 323 K. En la parte inferior se muestran los mismos ángulos característicos de las estructuras secundarias A y B-ADN [Olson].

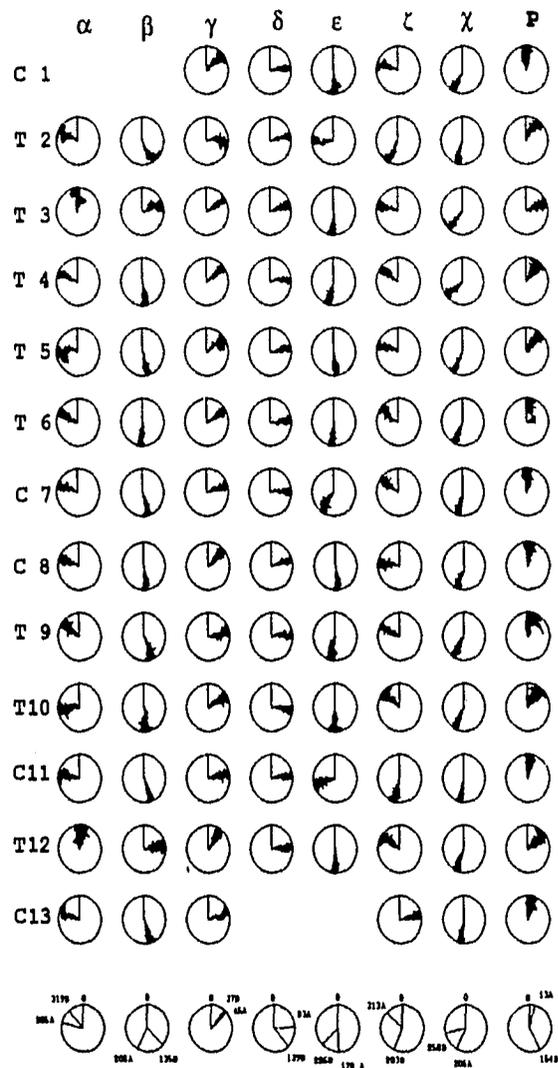


Figura 4.14: Ángulos de torsión y pseudorotación obtenidos para la cadena 3 a 323 K. En la parte inferior se muestran los mismos ángulos característicos de las estructuras secundarias A y B-ADN [Olson].

Nucleotidos	Energía Total (eV)
a) $T_{13}A_5T_3$	-10419.33
b) $T_{12}A_6T_4$	-10420.64
c) $C_8G_{10}C_8^+$	-10232.51
d) $T_7A_{11}T_9$	-10411.19

Tabla 4.5: Energía total de los nucleótidos seleccionados, a y b tienen como vecinos bases T:A:T, d tiene como vecino c una base C:G:C.

nuevamente se observa una dependencia con respecto a la secuencia del ADN.

Con el fin de estudiar la interacción intermolecular electrónica los puentes de hidrógeno tipo *Hoogsteen* que unen al oligo, se seleccionaron las bases a) y c) anteriores, pero ahora con la configuración a 10.6 ps obtenida de la dinámica a 293 K. Esto se hizo tomando en cuenta que los ángulos de torsión no tienen variaciones angulares muy fuertes debido a lo cual puede ser considerada, a *grosso modo*, como una estructura promedio. En este caso también los cálculos fueron a nivel semiempírico y se hicieron de la siguiente manera: calculando la energía total de las bases de *triplex* (T:A:T y C:G:C⁺), la energía total de las bases de *duplex* unidas con los puentes *Watson-Crick* (T:A y C:G) y por último las energías del nucleótido que corresponde al oligo (T y C), con estas energías se hicieron las diferencias $E_{a(tat)} - (E_{a(ta)} + E_{a(t)})$ y $E_{c(cgc^+)} - (E_{c(cg)} + E_{c(c^+)})$, con el fin de obtener solamente la energía de interacción entre las bases de la *duplex* y el oligo, unión *Hoogsteen*, las cuales se muestran en la tabla 4.6.

Bases	Energía total (eV)
$E_{a(tat)}$	-9456.89
$E_{a(ta)}$	-6292.57
$E_{a(t)}$	-3164.24
$E_{a(tat)} - (E_{a(ta)} + E_{a(t)})$	-0.08 (-1.8 Kcal/mol)
$E_{c(cgc^+)}$	-10224.54
$E_{c(cg)}$	-6987.41
$E_{c(c^+)}$	-3237.32
$E_{c(cgc^+)} - (E_{c(cg)} + E_{c(c^+)})$	0.19 (4.4 Kcal/mol)

Tabla 4.6: Energías totales electrónicas de las bases de *triplex* ($E_{a(tat)}$, $E_{c(cgc^+)}$), *duplex* ($E_{a(ta)}$, $E_{c(cg)}$) y oligo ($E_{a(t)}$, $E_{c(c^+)}$) con las diferencias que representan la energía intermolecular electrónica, dadas en eV y Kcal/mol, para comparar con lo obtenido en la sección de parámetros intermoleculares con el potencial de AMBER.

La densidad electrónica que rodea a las moléculas es una función tridimensional

que en general converge muy rápidamente a cero al alejarse de sus núcleos, por lo que la gráfica de esta densidad para valores mayores de un cierto número (muy pequeño), nos da la idea del concepto macroscópico que se tiene del cuerpo ó forma molecular, esta gráfica es útil ya que en ella se reflejan algunas propiedades tanto físicas como químicas de la molécula[Lipkowitz]. En la densidad electrónica graficada para las uniones triples T:A:T y C:G:C⁺ (fig. 4.15), se puede observar que las bases se encuentran completamente sombreadas y que hay regiones cuyo sombreado se une con la base contigua, reflejándose así las posiciones de los puentes de hidrógeno, que las mantienen unidas, ver sección 2.2. Es notorio que la unión triple C:G:C⁺ esta rodeada de una mayor densidad electrónica, lo cual debe estar relacionado con la mayor energía de interacción que esta unión tiene respecto de T:A:T (tabla 4.6). Relacionado con esto, en estudios de estabilidad experimentales se reporta que la estructura de *triplex* aquí estudiada se desnaturaliza a 37°C (se despega el oligo), sin embargo, cuando se reemplaza una unión triple T:A:T por un C:G:C⁺ en el centro de la secuencia, la estructura *triplex* se desnaturaliza a 38°C, asociando este resultado con una mayor estabilidad de la estructura [Mergny,et al,1991], siendo una posible causa de esto, la mayor interacción intermolecular electrónica del nucleótido C:G:C⁺. Análogamente, este resultado también puede tener relación con las configuraciones que adopta el azúcar cuando se tiene un vecino C:G:C⁺ en la secuencia, ya que al tener más energía de interacción es posible "brincar" la barrera de potencial y pasar a una configuración B-ADN (fig. 4.11).

4.2.4 Estructura del Solvente

Las propiedades del solvente que rodea el ADN es un factor que influye en la forma de su estructura secundaria de manera recíproca (sección 2.2). Se ha observado experimentalmente que la facilidad de formar y estabilizar moléculas *triplex* depende en gran medida de las condiciones en que se encuentre el solvente [Mergny,et al,1991]. Con esto en mente, se considero importante estudiar las características del solvente en la simulación con ayuda de la Función de Distribución Radial (FDR) ó de pares $g_{ab}(r)$. Esta función describe la densidad de probabilidad de encontrar un átomo tipo b a una distancia entre r y $r + dr$ de un átomo tipo a como una función de r y se calcula mediante la siguiente expresión,

$$g_{ab}(r) = \frac{d(N_{ab}(r))}{\rho_{ab}dV(r)},$$

1) 1000

2) 1000

1.5

3)

4)

5)

6)

7) 1000

a)



1.5

a)



b)



b)



1000x

Micrograph showing a cluster of small, dark, irregularly shaped particles.

donde $d\langle N_{ab}(r) \rangle$ es el número promedio de pares a - b en el diferencial de volumen $dV(r)$ dado entre r y $r + dr$ de un átomo a ó b , por lo cual $\frac{d\langle N_{ab}(r) \rangle}{dV(r)}$ es la densidad local promedio de átomos tipo a y b . En esta FDR, la densidad local se ha dividido por la densidad de átomos a y b en bulto ρ_{ab} , con objeto de que la función se normalice a largas distancias, en donde prácticamente ya no hay interacción entre los átomos. La FDR representa como los átomos de cierto tipo se organizan alrededor de otro tipo de átomo seleccionado (ó de ellos mismos). Si en las moléculas se encuentran átomos del mismo tipo, es posible hacer una partición entre los átomos intermoleculares e intramoleculares. La FDR depende de la densidad y de la temperatura, por ello su forma puede caracterizar al estado sólido, líquido ó gaseoso en que se encuentre el sistema. Por ejemplo, para sistemas cuyos átomos están fijos con posiciones bien definidas como en cristales, $g_{aa}(r)$ tiene una forma de funciones delta (ó funciones gaussianas pronunciadas, debido a las vibraciones del sistema). Para gases, por no tener una estructura definida, el comportamiento es completamente diferente, en este caso $g_{aa}(r)$ está dada aproximadamente por $\exp[-u(r)/KT]$, donde $u(r)$ es la función de energía potencial por pares [McQuarrie]. En los líquidos y sólidos amorfos el comportamiento es intermedio entre un cristal y un gas (fig. 4.16); los líquidos exhiben un comportamiento con un orden similar a un cristal a corto rango y a mayores distancias un desorden similar al de un gas. Los sólidos amorfos presentan el mismo comportamiento que el de un líquido pero un poco más enfático, incluso se observa a mayor rango [Haile]. Usando técnicas experimentales, es posible calcular la FDR mediante la transformación inversa de Fourier del factor de estructura $S(\omega)$ el cual se determina, dependiendo del sistema, mediante difracción de electrones, rayos X ó neutrones ([Elliott], p55-67).

El análisis de la estructura del solvente lo hice calculando FDR con el programa *decipher* manejado por *Insight*, promediando en los últimos 4ps (32.2-36.2 ps) de la dinámica a 293 K. Durante este intervalo de tiempo la energía potencial del sistema se mantiene aproximadamente constante y de esta forma es factible promediar sobre configuraciones que estén cerca del equilibrio termodinámico (fig. 4.4). Las diferencias más importantes observadas en la FDR obtenida entre los átomos de oxígeno de las moléculas de agua $g_{OO}(r)$ de esta simulación y una simulación de agua pura⁴ (fig. 4.17), son el incremento de la amplitud de las oscilaciones, la aparición de estas a mayores distancias, así como de la aparición del doble pico de la segunda oscilación. Aunque por otra parte, la FDR de la

⁴Esta simulación la hice bajo las mismas condiciones que la de 293 K pero sin ADN, sin contraiones y en una celda unitaria de $20 \times 20 \text{ \AA}^2$.

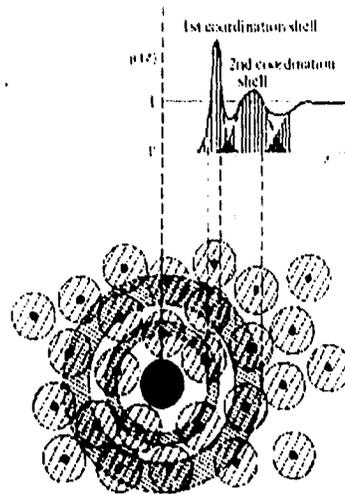


Figura 4.16: Ilustración esquemática de un arreglo de átomos correspondiente a un sólido amorfo (tomado de [Elliott]). Puede verse que la estructura de dicho arreglo se refleja en la forma de la FDR.

simulación de agua pura es muy similar a la obtenida experimentalmente para agua pura en estado líquido [Soper, et al, 1986], en donde solo se encuentra el primer pico un poco mayor y un desfase (menor de 1 Å) apartir de la segunda oscilación. Las causas de estas diferencias entre el agua pura son difíciles de precisar ya que pueden ser debidas a:

1. Las FDR se obtienen a diferentes temperaturas, con una diferencia de 5 K (que es poco significativa en realidad debido a que esto solo causa pequeños cambios en la amplitud [Postrino, et al, 1994]).
2. A que los parámetros de AMBER fueron ajustados para simular sistemas que contengan moléculas biológicas y no agua pura.
3. Deficiencias del modelo usado para generar el campo de fuerzas (AMBER), ecuación (3.9).
4. A dificultades inherentes de la técnica experimental de resolver distancias medias y largas [Trokhymchuck, et al, 1995].

Salvo las diferencias mencionadas, la estructura del solvente de la simulación de agua pura es parecida a la experimental. Asumiendo que el estado de la simulación de agua pura es líquido y tomando en cuenta lo mencionado sobre las distribuciones radiales, puedo decir que el estado del agua de la simulación con ADN y contraiones parecería relacionarse con un sólido amorfo. Debido a que las simulaciones se hicieron bajo las mismas condiciones, la única causa aparente de que el solvente adquiriera esta estructura es la presencia del ADN y los contraiones.

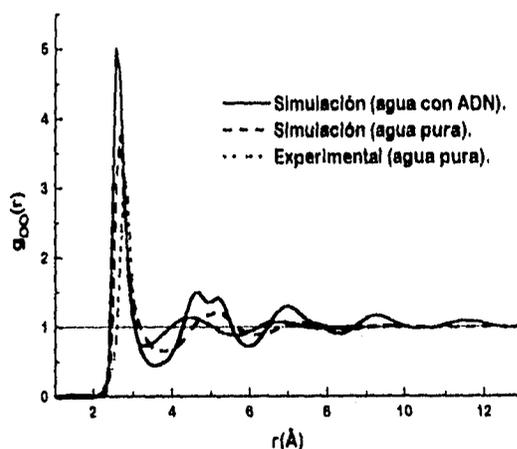


Figura 4.17: FDR intermolecular entre los O de las moléculas de agua. Para las simulaciones a 293 K (Obtenidos con *decipher*). El experimental a 298 K (tomado de [Soper86]).

Considerando que las interacciones a grandes distancias son principalmente ocasionadas por el campo eléctrico producido por las cargas del ADN y los contraiones, es probable que las moléculas de agua al polarizarse disminuyan su movilidad y correspondiente con esto un aumento de su viscosidad, fenómeno al que se le llama **efecto de electroviscosidad** ([Richards], p 235). Ocasionado por esta disminución de la movilidad de las moléculas de agua es factible que los tiempos de vida media de los puentes de hidrógeno formados entre las moléculas de agua se incrementen⁵, análogamente a lo que sucede en

⁵En el agua las uniones $O-H$ actúan como donadores y aceptores a la vez, formandose el puente $O-H \cdots O$.

los compuestos llamados *clathrate*⁶ [Davidson]. Por lo cual, si los puentes del hidrógeno formados entre las moléculas de agua que rodean al ADN tienen tiempos de vida media mayores a los tiempos manejados en esta simulación (10^{-12} seg), no es posible ver que ocurra el rompimiento de estos puentes y por lo tanto la estructura del agua en esta escala de tiempo sea muy semejante a la de un sólido en vez de un líquido⁷. Relacionado con este punto, en la literatura encontré reportes tanto de experimentos como de simulaciones en donde se observa que el comportamiento del solvente que rodea a las moléculas de ADN, es análogo al obtenido en esta simulación (llamándolo *freezing behavior*, determinado experimentalmente con infrarrojo [Falk, et al, 1970] y por sus números de hidratación en simulaciones [Forester, et al, 1991]). Es razonable pensar que la causa del doble pico observado en la segunda oscilación de $g_{OO}(r)$ de la simulación con ADN y contraiones, se deba a este comportamiento viscoso, ya a que el acomodo de las moléculas de agua en estas condiciones debe ser distinto que cuando la viscosidad es baja (agua pura). En general el acomodo del que hablo debe ser ocasionado por la forma de la molécula ya que esta no es esférica (ver fig. 4.16), lo cual se acentúa más al no tener libertad de rotar libremente a causa de la disminución de su movilidad ocasionada por la orientación de los dipolos de la molécula de agua. Debo mencionar que un doble pico similar aparece también en la segunda oscilación de estructuras vídrias determinadas experimentalmente [Elliott].

De las gráficas $g_{PO}(r)$ y $g_{NaO}(r)$ calculadas (figs. 4.18 y 4.19) puedo decir por sus oscilaciones que tanto en el ADN como en los contraiones se forman capas de agua, observándose también el efecto de **exclusión de volumen**, dado que las oscilaciones se mantienen por debajo de la normalización que se hace para el solvente en bulto a 20 Å. Este efecto es ocasionado por la magnitud de la molécula de ADN, ya que debido a su presencia disminuye la cantidad de moléculas de agua por unidad de volumen, como puede observarse en los esquemas de dichas figuras. El efecto de exclusión de volumen es claro para los grupos fosfato que forman parte de la molécula de ADN, sin embargo este también se observa en la distribución de agua alrededor de los contraiones, lo cual significa que los contraiones se encuentran relativamente cerca de la molécula de ADN. En estas gráficas también se observa un doble pico como el que se hablaba en la $g_{OO}(r)$, en la tercer oscilación

⁶El agua se comporta como un *clathrate* bajo ciertas condiciones y sus puentes de hidrógeno tienen tiempos de vida media el orden de 10^{-10} seg ([Chang], p 500) y también se sabe que bajo esas condiciones presenta estructuras semejantes a las del hielo (*ice-like*) [Soper, et al, 1994].

⁷Un ejemplo muy claro que ocurre en la naturaleza es el del vidrio, el cual en nuestra escala de tiempo tiene una apariencia sólida, sin embargo en periodos muy largos este se comporta como un fluido.

para $g_{NaO}(r)$ y en la segunda para $g_{PO}(r)$, aunque en este caso se piensa que puede ser un efecto del movimiento de los contraiones ([Blackburn], p 299). Un punto sobre estas gráficas que me parece interesante remarcar relacionado con la viscosidad, es la polarización de las moléculas de agua alrededor de los grupos fosfato y de los contraiones, que se ve claramente debido a que el primer pico de la gráfica $g_{PH}(r)$ se encuentra un poco más cerca que en $g_{PO}(r)$ (fig. 4.18) y en $g_{NaH}(r)$ un poco más lejos que $g_{NaO}(r)$ (fig. 4.19).

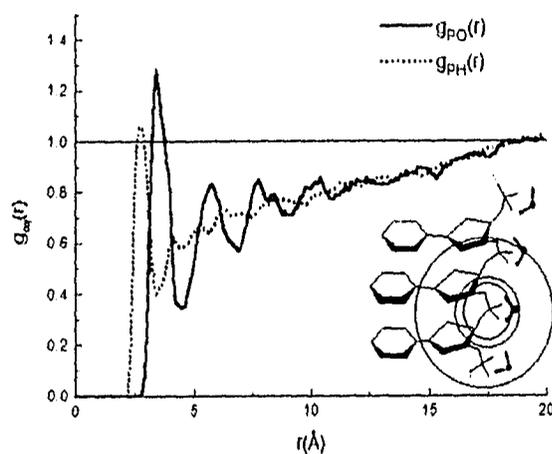


Figura 4.18: FDR intermolecular de la simulación a 293 K entre los átomos P-O y P-H de los fosfatos en las cadenas y moléculas de agua. Obtenidos con *decipher*.

La distribución de los contraiones alrededor de los grupos fosfato del ADN $g_{PNa}(r)$ y la de los contraiones alrededor de ellos mismos $g_{PNa}(r)$ (fig. 4.20) son muy similares, tienen picos muy altos en varias posiciones y mucho *ruido*, esto último creo que se debe a que el número de iones es pequeño comparado con el número de átomos del sistema. Sin embargo, los picos de la gráfica se deben a que los contraiones se mantienen alrededor del ADN rodeando a los grupos fosfato, de la misma manera que en las posiciones encontradas en la sección 4.1.2, a esto se le conoce con el nombre de **condensación de iones** y concuerda con la exclusión de volumen mencionada anteriormente [McCammon].

En general las estructuras A-ADN se forman experimentalmente en condiciones de baja humedad y altas concentraciones de sal y las B-ADN con alta humedad y bajas

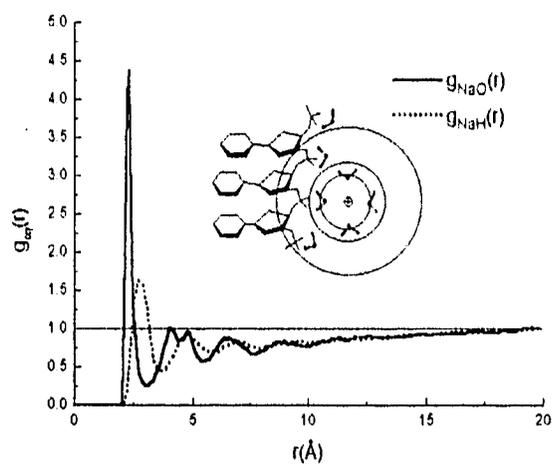


Figura 4.19: FDR intermolecular de la simulación a 293 K entre los contraiones (Na), O y H de las moléculas de agua. Obtenidos con *decipher*.

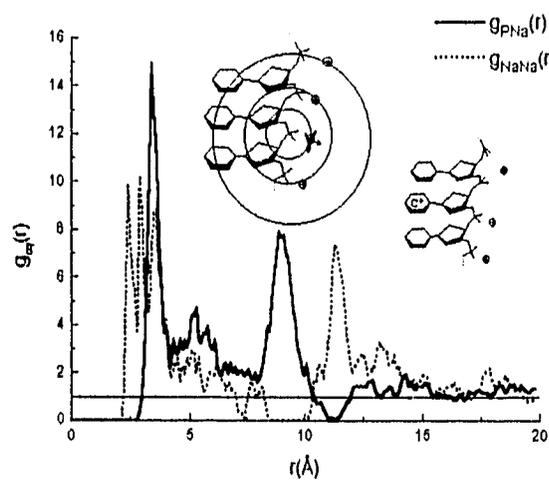


Figura 4.20: FDR intermolecular de la simulación a 293 K entre los átomos P de las cadenas y los contraiones (Na). Obtenidos con *decipher*.

concentraciones de sal. Por lo que es posible que bajo las condiciones en que se encuentra esta simulación en tiempos posteriores (del orden de segundos [McCammon]) se observe una transición global a B-ADN, lo cual corresponde con las transiciones de la configuración del azúcar y de algunos ángulos de torsión observados en la sección 4.2.2. En los estudios de estabilidad experimentales, se reporta que con el incremento de la cantidad de sales (iones) en la solución no se observa la histéresis de la absorción óptica ocasionada por la desnaturalización y renaturalización incompleta de la estructura secundaria del ADN-*triple* [Mergny, et al, 1991], facilitándose con ello la formación de las estructuras. Correspondiente con esta observación existe el hecho de que el incremento de la concentración de sales también disminuye el efecto de electroviscosidad ([Richards], p 235), una posible causa de ello podría ser que la solución al ser menos viscosa permite mayor movilidad en las cadenas haciendo más fácil el reconocimiento entre las bases. Al respecto, se sabe que la acción de los campos eléctricos facilita la movilidad de las cadenas de ADN y esto se emplea en la **electroforesis** la cual es una técnica experimental usada para filtrar dichas cadenas ([Richards] p 199, 232).

Capítulo 5

Conclusiones

Sobre los programas de cómputo desarrollados y paquetes de cómputo empleados

- Se desarrollaron e implementaron los siguientes programas y utilerías que en conjunto facilitan el desarrollo de la simulación y el análisis de los resultados. La mayoría de ellos pueden ser usados en estudios similares.
 - Utilerías Unix para el intercambio de información entre diferentes formatos de los paquetes y programas (e.g. cargas y coordenadas atómicas).
 - Programas Fortran para encontrar la posición de contraiones, usando diferentes métodos (e.g. Steepst, gradientes, criterio del Hessiano y valor mínimo, donde este último resultó ser más eficiente).
 - Macros BCL que trabajan en conjunto con programas Fortran y C, para extraer información de las dinámicas, calcular y graficar (e.g. ángulos de torsión y energías intermoleculares).
- Se usaron 2 métodos cuánticos para el cálculo de cargas atómicas del nucleótido C⁺ (Mulliken y ajuste al potencial electrostático) y se eligió el segundo con el único objeto de mantener congruencia con el método utilizado en AMBER para este cálculo.
- La comparación cualitativa de dos simulaciones usando los ensambles NPT y NVT, me condujeron a la elección del NPT por considerarlo más representativo de lo que ocurre naturalmente (sección 4.1.4).

- La forma de las gráficas de energía potencial vs. tiempo (fig. 4.4), sugiere un incremento en el tiempo de inicialización para alcanzar la estabilidad del sistema.
- El procedimiento para la obtención de una configuración inicial, lo considero razonable para iniciar la simulación a partir de la configuración de mínima energía encontrada, ya que se partió de información cristalográfica y se tuvo el cuidado de no deformar la estructura debido al solvente y a los contraiones (sección 4.1.3).
- De los paquetes de visualización utilizados debo mencionar que las representaciones gráficas moleculares son muy ilustrativas de los fenómenos cuánticos y de las características estructurales de las moléculas estudiadas. Por lo que este tipo de estudios me fué muy didáctico para entender muchas propiedades moleculares físicas, químicas y biológicas.

Sobre al análisis e interpretación

Se hicieron algunas observaciones que confirman algunos de los problemas que se encuentran para el desarrollo de fármacos mencionados en la introducción, no obstante obtuvimos elementos que se relacionan con observaciones experimentales y que pueden ser útiles en los siguientes aspectos:

- Se pueden observar indicios de una desestabilización local que en las estructuras *triplex* añade la presencia del oligo al *duplex* a 293 K, esto mediante la comparación de los parámetros intermoleculares de las simulaciones a 293 K y 323 K, lo cual me parece razonable ya que concuerdan cualitativamente con la interpretación de los experimentos de estabilidad determinados mediante absorción óptica [Mergny, et al, 1991].
- A partir del análisis hecho con los ángulos de torsión y pseudorotación, se observó en algunos casos que la estructura del ADN-triplex depende de su secuencia, lo cual está relacionado con observaciones experimentales hechas con RMN.
- En los estudios de estructura electrónica hechos a las uniones tipo *Hoogsteen* (sección 4.2.3), también se observó la dependencia de la estructura con la secuencia.
- Con las FDR hechas para estudiar la distribución del solvente que rodea el ADN, se obtuvieron resultados que pueden estar relacionados con observaciones experimentales sobre su viscosidad, esto se hizo observando la forma en que la presencia del ADN y los iones modifican la estructura del agua (sección 4.2.4).

Sobre el modelado molecular

- La descripción de un sistema biológico de este tipo pese a todos los recursos con que se contó para el desarrollo de este estudio, es una tarea bastante compleja y difícil de llevar al cabo.
- Dentro de las aproximaciones mencionadas en el capítulo 3, puedo decir que el modelo usado representa de forma adecuada las propiedades moleculares del sistema estudiado, debido a que muchos de los resultados obtenidos concuerdan cualitativamente con observaciones experimentales de estos sistemas macromoleculares.
- Como mencioné en la sección 4.2.4, las interacciones electrostáticas están muy relacionadas con la viscosidad obtenida en las FDR del solvente. Esta información creo que es importante tomarse en cuenta para futuros estudios, ya que es posible controlar las características del solvente (e.g. viscosidad) ya sea agregando iones en el medio y/o ajustando una constante dieléctrica (en esta simulación $\epsilon = 1$), de tal forma que se pueda disminuir o aumentar las interacciones electrostáticas [Yang, et al, 1995]. O bien, dada la importancia del solvente, buscar nuevas representaciones (modelos) [Caldwell, et al, 1990] y [Dang, et al, 1991], ya que de él dependen tanto forma de la estructura del ADN como su estabilidad misma.
- Finalmente, es lamentable que un solo estudio no sea suficiente para extraer toda la información deseada, a lo largo de la elaboración la presente tesis he pensado en la posibilidad de dar continuidad a este trabajo con algunos de los siguientes puntos:
 - Analizar con diferentes secuencias el polimorfismo estructural del ADN dependiente de la secuencia y el solvente, el uso de *mismatch* en el oligo, como también probar la estabilidad de nuevos nucleótidos.
 - Trabajar con estudios de estructura electrónica al respecto de las condiciones de ángulo y distancia para la formación de puentes de hidrógeno, así como de la dependencia secuencial de las configuraciones del azúcar.
 - Estudiar el comportamiento polielectrolítico para obtener mejores representaciones, dentro del cual es importante: el cálculo de cargas parciales, incluir una constante dieléctrica ó usar nuevos campos de fuerza.

Capítulo 6

Apendices

6.1 Método de Hartee-Fock

(Tomado de [Weissbluth] y [Levine])

Para poder efectuar los estudios de estructura electrónica molecular, más allá de moléculas homo-nucleares, es necesario hacer una gran cantidad de cálculos que van acompañados de aproximaciones y métodos que nos ayudan a resolver la ecuación de Shrödinger (3.4.)

Propuesto intuitivamente por D. R. Hartee (1928) con argumentos físicos y probado mediante el método variacional por V. A. Fock, que introduce antisimetría con un determinante de Slater, (1930) [Levine], el método Hartree-Fock (H-F) ha demostrado ser apropiado para el estudio de sistemas nucleares, atómicos, moleculares y sólidos.

La solución del problema molecular es muy similar al problema atómico, pero es mucho más compleja, debido a la ausencia de simetrías rotacionales y el aumento en el número de núcleos. Para ello, se postula el hecho de poder calcular los orbitales moleculares (OM) como una combinación lineal de orbitales atómicos (CLOA) [Weissbluth]. De esta forma el OM $\psi(\mathbf{r})$ es escrito como:

$$\psi(\mathbf{r}) = \sum_{\mu=1}^K C_{\mu} \phi_{\mu}(\mathbf{r}), \quad (6.1)$$

En donde $\phi_{\mu}(\mathbf{r})$ son orbitales atómicos. Este método fué propuesto por Roothaan y simplifica enormemente el problema electrónico molecular ([Weissbluth] p. 556). Los OM deben cumplir con las siguientes características:

1. El OM es una eigenfunción de un Hamiltoniano unielectrónico.
2. Un OM puede extenderse sobre cualquier número de átomos en una molécula.
3. No más de dos electrones pueden ocupar el mismo orbital molecular y estos deben tener espín opuesto, principio de Pauli.
4. Definición: un orbital $\psi(\mathbf{r})$ multiplicado por una función de espín $\alpha(\omega)$ o $\beta(\omega)$ es un espín-orbital molecular $\chi(\mathbf{x}) = \psi(\mathbf{r})\alpha(\omega)$ o $\chi(\mathbf{x}) = \psi(\mathbf{r})\beta(\omega)$.

Dentro del esquema conocido como H-F restringido y limitándonos a la discusión sobre moléculas de capa cerrada, la función de onda antisimétrica, que puede ser usada para describir el estado base de un sistema de $2N$ electrones, es un solo **determinante de Slater**, el cual tiene la siguiente forma:

$$\Psi = \frac{1}{\sqrt{(2N)!}} \begin{vmatrix} \chi_1(\mathbf{x}_1) & \chi_2(\mathbf{x}_1) & \cdots & \chi_N(\mathbf{x}_1) \\ \chi_1(\mathbf{x}_2) & \chi_2(\mathbf{x}_2) & \cdots & \chi_N(\mathbf{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \chi_1(\mathbf{x}_{2N}) & \chi_2(\mathbf{x}_{2N}) & \cdots & \chi_N(\mathbf{x}_{2N}) \end{vmatrix} = |\chi_1(\mathbf{x}_1)\chi_2(\mathbf{x}_2)\cdots\chi_{2N}(\mathbf{x}_{2N})|, \quad (6.2)$$

De acuerdo en el principio variacional, tomamos los espín-orbitales moleculares que minimizan la energía electrónica molecular $E = \langle \Psi | H | \Psi \rangle$, es decir $\delta E = 0$, sujetos a la condición $\langle \Psi | \Psi \rangle = 1$. Aplicando el método de multiplicadores de Lagrange se obtiene,

$$\delta \{ \langle \Psi | H - \lambda | \Psi \rangle \} = 0,$$

de donde

$$(H - \lambda) | \Psi \rangle = 0, \quad \langle \Psi | (H - \lambda) = 0,$$

identificando estas dos ecuaciones con la ecuación de Schrödinger (3.4), podemos asumir que λ sea el eigenvalor de la energía E . Tomando las variaciones correspondientes de $\delta \Psi$ podemos acotar a la energía E con el estado de mínima energía E_0 obteniendo,

$$E_0 \leq \langle \Psi_0 | H_e | \Psi_0 \rangle.$$

De esta forma y tomando en cuenta (6.1) se deriva la ecuación de Hartee-Fock:

$$f(i)\chi(\mathbf{x}_i) = \varepsilon\chi(\mathbf{x}_i). \quad (6.3)$$

donde $f(i)$ es un operador efectivo de un electrón, llamado operador de Fock

$$f(i) = -\frac{1}{2}\nabla_i^2 - \sum_{A=1}^M \frac{Z_A}{r_{iA}} + v^{HF}(i),$$

donde $v^{HF}(i)$ es el potencial promedio que experimenta el electrón i -ésimo debido a la presencia de los otros electrones. La esencia de la aproximación de Hartree-Fock es reemplazar el complicado problema multielectrónico por un problema unielectrónico en donde la repulsión electrón-electrón es tratada como un potencial promedio.

El potencial Hartree-Fock $v^{HF}(i)$, o el campo equivalente visto por el i -ésimo electrón, depende de los espín-orbitales de los otros electrones. Debido a ello la ecuación (6.3) es no-lineal y puede ser resuelta iterativamente.

El procedimiento para resolver la ecuación de Hartree-Fock es llamado método de campo autoconsistente (SCF, *self-consistent-field*), cuya idea principal se basa en tomar un conjunto de OM iniciales y con ellos calcular en potencial promedio $v^{HF}(i)$ que siente cada electrón y con esto resolver la ecuación de eigenvalores (6.3) para obtener un nuevo conjunto de OM. Usando estos nuevos OM para obtener nuevos campos y repetir el procedimiento hasta que se observen mínimas variaciones de la energía y los OM sean igual a las eigenfunciones, es decir, que el procedimiento sea autoconsistente.

6.2 Orden de enlace

(Tomado de [Ofenhardt] y [Szabo])

En el esquema descrito en la sección 6.1 sobre moléculas de capa cerrada, donde cada orbital molecular Ψ_a está doblemente ocupado, la probabilidad de encontrar un electrón cualquiera en un elemento de volumen es $\rho(\mathbf{r})d\mathbf{r}$, donde,

$$\rho(\mathbf{r}) = 2 \sum_{a=1}^N |\Psi_a(\mathbf{r})|^2,$$

es la densidad de carga. Mediante la expansión (6.1) para cada OM la densidad de carga se transforma en;

$$\begin{aligned} \rho(\mathbf{r}) &= 2 \sum_{a=1}^N \Psi_a^*(\mathbf{r})\Psi_a(\mathbf{r}) = \\ &= \sum_{\mu\nu} \left[2 \sum_{a=1}^N C_{\mu a} C_{\nu a}^* \right] \phi_\mu(\mathbf{r})\phi_\nu^*(\mathbf{r}) = \sum_{\mu\nu} P_{\mu\nu} \phi_\mu(\mathbf{r})\phi_\nu^*(\mathbf{r}), \end{aligned}$$

donde hemos definido la matriz de densidad ó matriz de orden de enlace.

$$P_{\mu\nu} = 2 \sum_{a=1}^N C_{\mu a} C_{\nu a}^*$$

donde μ, ν son los índices de los orbitales atómicos y 'a' el índice de los OM. Para $\mu \neq \nu$, los elementos de la matriz corresponden a la noción convencional de enlace entre los átomos μ y ν . Para $\mu = \nu$ nos representa la densidad total electrónica en el μ -ésimo átomo.

6.3 Análisis de población de Mulliken

(Tomado de [Offenhardt] y [Szabo])

Para modelar las propiedades electrostáticas de una molécula es común recurrir al concepto de carga atómica, ya que con estas cargas es posible representar las interacciones electrostáticas útiles en las simulaciones moleculares para representar propiedades fisicoquímicas [Bayly, et al, 1993-97]. Debido a que en mecánica cuántica no hay definido un operador asociado al concepto de carga atómica, R. S. Mulliken propuso en 1955 lo que en la actualidad conocemos como el **análisis de población de Mulliken**. Este análisis parte de que el número de electrones total en el sistema esta dado por:

$$2N = \sum_{\mu} \sum_{\nu} P_{\mu\nu} S_{\nu\mu} = \sum_{\mu} (\mathbf{PS})_{\mu\mu} = \text{tr} \mathbf{PS},$$

donde la matriz \mathbf{S} es conocida como la matriz de traslape y tiene como elementos:

$$S_{\nu\mu} = \int \phi_{\nu}^*(\mathbf{r}) \phi_{\mu}(\mathbf{r}) d\mathbf{r}.$$

Comúnmente se interpreta el término $(\mathbf{PS})_{\mu\mu}$ como el número de electrones asociados con el orbital atómico ϕ_{μ} . El número total de electrones asociados a un determinado átomo de una molécula se obtiene sumando $(\mathbf{PS})_{\mu\mu}$ sobre los orbitales atómicos centrados en ese átomo. La carga neta asociada con un átomo esta dada por.

$$q_A = Z_A - \sum_{\mu \in A} (\mathbf{PS})_{\mu\mu}. \quad (6.4)$$

En el contexto de simulaciones moleculares esta carga en algunas ocasiones se emplea como la carga parcial asociada al átomo A [Singh, et al, 1984].

6.4 Superficie molecular de potencial electrostático

(Tomado de [Offenhardt] y [Szabo])

Para alguna conformación particular de los núcleos atómicos de una molécula, basados en nuestro esquema Born-Oppenheimer, el potencial electrostático $V(\mathbf{r})$ producido en algún punto del espacio tendrá contribuciones tanto nucleares como electrónicas. De tal forma que,

$$V(\mathbf{r}) = \sum_A \frac{Z_A}{|\mathbf{r} - \mathbf{R}_A|} - \int \frac{\rho(\mathbf{r})}{|\mathbf{r} - \mathbf{r}_{electrónico}|} d\mathbf{r}_{electrónico}, \quad (6.5)$$

Para el cual pueden ajustarse cargas q_A asociadas al átomo A.

6.5 Método de dinámica molecular

(Tomado de [Haile] y [Discover])

El método de dinámica molecular se basa esencialmente en resolver numéricamente las ecuaciones clásicas de movimiento de un conjunto de interacciones de cada uno de los átomos, ec. (6.6) a partir de sus coordenadas y velocidades iniciales. Donde las coordenadas pueden ser obtenidas de datos cristalográficos, NMR, mecánica molecular o métodos cuánticos y las velocidades mediante una distribución de Maxwell-Boltzmann.

$$f(v)dv = \left(\frac{m}{2\pi kT}\right)^{\frac{3}{2}} e^{-mv^2/2kT} 4\pi v^2 dv.$$

La cual expresa la probabilidad de que una molécula de masa m tenga la velocidad v cuando está a la temperatura T . Las ecuaciones que deben resolverse forma un conjunto de ecuaciones diferenciales acopladas de segundo orden cuyo número es igual al número de grados de libertad de el sistema.

$$\mathbf{F}_i(t) = m_i \mathbf{a}_i(t) = m_i \frac{\partial^2 \mathbf{r}_i}{\partial t_i^2} = m_i \frac{\partial \mathbf{v}_i}{\partial t_i}, \quad i = 1, N \text{ (número de átomos)}, \quad (6.6)$$

Donde la fuerza que actúa sobre átomo i al tiempo t es calculada como el gradiente negativo de la función de energía potencial

$$\mathbf{F}_i = -\frac{\partial}{\partial \mathbf{r}_i} V(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N). \quad (6.7)$$

El tipo de algoritmo usado para la integración numérica de estas ecuaciones newtonianas depende del tipo de interacción entre los átomos del sistema.

Existe una gran cantidad de algoritmos que pueden ser usados para integrar las ecuaciones de movimiento (6.6). El algoritmo de Verlet (1967) es el más común, calcula las nuevas posiciones usando las fuerzas y las posiciones previas de los átomos. De esta manera, si \mathbf{r}_i es la posición del átomo i al tiempo t , esta puede ser calculada a $t + \Delta t$ mediante un desarrollo en serie de Taylor:

$$\mathbf{r}_i(t \pm \Delta t) = \mathbf{r}_i(t) \pm \dot{\mathbf{r}}_i(t)\Delta t + \ddot{\mathbf{r}}_i(t)\frac{(\Delta t)^2}{2!} \pm \frac{\partial^3 \mathbf{r}_i(t)}{\partial t^3} \frac{(\Delta t)^3}{3!} + \dots \quad (6.8)$$

Sumando estas expresiones, despejando y sustituyendo en (6.6), se obtiene:

$$\mathbf{r}_i(t + \Delta t) = -\mathbf{r}_i(t - \Delta t) + 2\mathbf{r}_i(t) + (\Delta t)^2 \frac{1}{m_i} \sum_{j \neq i} \mathbf{F}_{ij}(t) + O(\Delta t)^4, \quad (6.9)$$

La velocidad puede calcularse por la fórmula de distancia central

$$\mathbf{v}_i(t) = \frac{\mathbf{r}_i(t + \Delta t) - \mathbf{r}_i(t - \Delta t)}{2\Delta t} + O(\Delta t)^3, \quad (6.10)$$

Dada las velocidades, $\mathbf{v}_i(t)$, es posible calcular la temperatura del sistema al tiempo t de la ecuación.

$$T(t) = \frac{1}{(3N - n)k_B} \sum_{i=1}^N m_i |\mathbf{v}_i|^2, \quad (6.11)$$

Para que un estado termodinámico esté definido completamente es necesario incluir a la presión. Aunque la presión esta determinada por una variable escalar, esta puede ser representada por un tensor en una forma general ([McQuarrie] 1976):

$$P = \begin{bmatrix} P_{xx} & P_{xy} & P_{xz} \\ P_{yx} & P_{yy} & P_{yz} \\ P_{zx} & P_{zy} & P_{zz} \end{bmatrix}.$$

En donde cada elemento del tensor actúa en la superficie de un volumen cúbico infinitesimal.

6.5.1 Ensambls NVE, NVT y NPT.

En una simulación con dinámica molecular es posible manejar los tres principales tipos de ensambles, conocidos en la mecánica estadística; NVE, NVT y NPT. Dependiendo del tipo de sistema y de las condiciones en las que se encuentre el sistema que se desee simular es preferible utilizar algún tipo de ensamble específico.

NVE Microcanónico.

Este se obtiene al resolver directamente las ecuaciones de Newton con el algoritmo de *Verlet's Leap frog*, sin ningún ajuste de temperatura ni presión durante la simulación. Este no es recomendable durante la primera parte de la dinámica (inicialización) debido a que sin el flujo de energía obtenido por el contacto con un baño térmico o reescalando las velocidades, la temperatura deseada no será obtenida. Sin embargo durante la fase de recolección de datos (resume), si se está interesado en explorar la superficie de energía constante del espacio de conformaciones, o por alguna otra razón en la que no se desee las perturbaciones introducidas por el acoplamiento de un baño térmico o baño de presión.

NVT Canónico.

Para este ensamble es necesario controlar la temperatura, reescalando las velocidades (durante el equilibrio) y acoplando un baño térmico durante la dinámica. El volumen se mantiene constante a través de la corrida. Esta es una elección apropiada cuando se busca la conformación de la molécula en vacío y sin condiciones periódicas, porque sin condiciones periódicas el volumen, presión y densidad no son definidas. Aunque si se usan condiciones periódicas, si la presión no es un factor significativo, el NVT tiene la ventaja de pequeñas perturbaciones en la trayectoria debido a la ausencia del acoplamiento a un "baño de presión".

NPT.

Controla el valor de la temperatura y la presión, permitiendo que el volumen varíe, mediante el ajuste de los vectores de la celda unitaria. Este ensamble se elige en casos donde la presión correcta, volumen y densidad, son factores importantes en los cálculos. Este ensamble se puede usar durante el equilibrio para adquirir la temperatura y presión deseada antes de cambiar a NVE ó NVT cuando la recolección de datos inicia [Discover].

6.6 Programa para calcular la posición de contraiones

```

C *****
  PROGRAM SELF15
C A.A.HUERTA, DICIEMBRE 1994.
C *****
C DECLARACION DE VARIABLES USADAS
C *****
  parameter (natms=1500,nmin=40,nth=157,nz=80)
  common /coord/ a(natms),b(natms),c(natms),q(natms)
  common /poten/ V(nth,nz),Vm(nmin)
  common /ctes/ iatms,carg,cmin,cmax,dz,dth,r
  dimension xm(nmin),ym(nmin),zm(nmin),qm(nmin)
  real carga,Pi
  real xmi,ymi,zmi,dif
  integer flag
  Pi=3.14159265
  CALL LEER(carga)
  dz = (cmax-cmin)/nz
  dth = 2*Pi/nth
C *****
C calcula el potencial electrostatico producido por todas las
C cargas parciales.
C *****
  r=15.0
  ih=1
  CALL POTENCIAL(ih,im,km)
C *****
C Asigna a la posicion de minimo pot. un contraion
C *****
1  zm(ih) = cmin + dz*km
   xm(ih) = r*cos(dth*im)
   ym(ih) = r*sin(dth*im)

```

```

qm(ih) = 1.000
write (*,35) xm(ih),ym(ih),zm(ih),qm(ih)
C *****
C Calcula el nuevo potencial debido al contraion
C *****
CALL POTENCIAL1(ih,xm(ih),ym(ih),zm(ih),qm(ih),im,km)
if (ih.eq.carg) goto 2
ih=ih+1
goto 1
2 continue
C *****
C Rutina autoconsistente
C *****
5 flag=0
do ih=1,carg
qm(ih)=-qm(ih)
xmi=xm(ih)
ymi=ym(ih)
zmi=zm(ih)
CALL POTENCIAL1(ih,xm(ih),ym(ih),zm(ih),qm(ih),im,km)
zm(ih) = cmin + dz*km
xm(ih) = r*cos(dth*im)
ym(ih) = r*sin(dth*im)
qm(ih) = 1.000
CALL POTENCIAL1(ih,xm(ih),ym(ih),zm(ih),qm(ih),im,km)
write (*,35) xm(ih),ym(ih),zm(ih),qm(ih)
dif=sqrt( (xmi-xm(ih))*(xmi-xm(ih)) + (ymi-ym(ih))*(ymi-ym(ih)) +
n (zmi-zm(ih))*(zmi-zm(ih)) )
if (dif.ge.1.0) flag=1
enddo
if (flag.eq.1) goto5
STOP
35 format(3f15.9,f7.2)

```

```

END
C *****
SUBROUTINE LEER(carga)
C lee x,y,z de las cargas
C *****
parameter (natms=1500,nmin=40,nth=157,nz=80)
common /coord/ a(natms),b(natms),c(natms),q(natms)
common /ctes/ iatms,carg,cmin,cmax,dz,dth,r
real carga
open (10, file='tri.dat')
read (10,*) iatms
read (10,*) (a(i),b(i),c(i),q(i),i=1,iatms)
close (10)
C *****
C suma la carga total
C busca el cmin y cmax, en el eje de la molecula (z)
C *****
carga=0
cmax=c(1)
cmin=c(1)
do 10 i=1,iatms
carga = carga + q(i)
if (c(i) .GE. cmax) cmax=c(i)
if (c(i) .LE. cmin) cmin=c(i)
10 continue
carg=-int(carga)
write (*,*) carga,cmin,cmax,iatms
RETURN
END
C *****
SUBROUTINE POTENCIAL(ih,im,km)
C *****
parameter (natms=1500,nmin=40,nth=157,nz=80)

```

```

common /coord/ a(natms),b(natms),c(natms),q(natms)
common /poten/ V(nth,nz),Vm(nmin)
common /ctes/ iatms,carg,cmin,cmax,dz,dth,r
do 40 k = 1,nz
  z = cmin+dz*(k)
  do 30 i = 1,nth
    x = r*cos(dth*i)
    y = r*sin(dth*i)
    do 20 j = 1,iatms
      dist = sqrt( (x-a(j))*(x-a(j)) + (y-b(j))*(y-b(j)) +
n      (z-c(j))*(z-c(j)) )
      if (dist .LE. 0.0001) goto 30
      V(i,k) = q(j)/dist + V(i,k)
20 continue
C *****
C Determina cual es el valor minimo del potencial
C *****
  if ( i.eq.1.and.k.eq.1) Vm(ih)=V(i,k)
  if ( V(i,k) .le. Vm(ih) ) then
    Vm(ih) = V(i,k)
    im = i
    km = k
  endif
30 continue
40 continue
  RETURN
  END
C *****
  SUBROUTINE POTENCIAL1(ih,xmih,ymih,zmih,qmih,im,km)
C *****
  parameter (natms=1500,nmin=40,nth=157,nz=80)
  common /coord/ a(natms),b(natms),c(natms),q(natms)
  common /poten/ V(nth,nz),Vm(nmin)

```

```

common /ctes/ iatms,carg,cmin,cmax,dz,dth,r
do 40 k = 1,nz
z = cmin + dz*(k)
do 30 i = 1,nth
x = r*cos(dth*i)
y = r*sin(dth*i)
dist = sqrt( (x-xmih)*(x-xmih) + (y-ymih)*(y-ymih) +
n (z-zmih)*(z-zmih) )
c jac if (dist .LE. 0.0001) goto 30
if (dist .LE. 0.0001) dist=0.0001
V(i,k) = qmih/dist + V(i,k)
C *****
C Determina cual es el valor minimo del potencial
C *****
if ( i.eq.1.and.k.eq.1) Vm(ih)=V(i,k)
if( V(i,k) .le. Vm(ih) ) then
Vm(ih) = V(i,k)
im = i
km = k
endif
30 continue
40 continue
c write (*,*) 'e',ih,im,km
RETURN
END

```

Bibliografía

- [Blackburn] Blackburn G. M. and Gait M. J. , *Nucleic Acids in Chemistry and Biology*, Oxford university press, 1990.
- [Chang] Raymond Chang, *Physical Chemistry, (with applications to biological Systems*. Macmillan Publishing, NY, 1990.
- [Cohen-T] Cohen-Tannoudji C. Francklaloé B., *Quantum Mechanics*, A Wiley-Interscience pub., 1977.
- [Davidson] Davidson D.W, *Water:A Comprehensive Treatise* Plenum, New York, 1973.
- [Elliott] Elliott S.R., *Physics of amorphous materials* Longman Scientific and Technical, England, 1989.
- [Galli] Galli G. and Parrinello M., *Computer Simulation in Material Science*, M. Meyer and V. Pontikis, Ed. Kluwer, Dordrecht, 1991.
- [Haile] Haile M. J. , *Molecular Dynamics Simulation. Elementary Methods*, John wiley & Sons, inc., 1992.
- [Heermans] Heermans J., *Notas del Curso sobre dinámica Molecular*, UNAM, Nov 15-19, 1993.
- [Holde, van] Holde, van K. E., *Bioquímica Física*, EXEDRA, Ed. Alhambra.
- [Hostetter] Hostetter G.H., Santana M.S., *Analytical, numerical and computational methods for science and engineering*, Prentice Hall.

- [Levine] Levine I.N., *Quantum Chemistry*, 2a.ed, Allyn and Bacon, 1974. y 4a.ed, Prentice Hall, 1991.
- [Lipkowitz] Lipkowitz and Boyd *Review in Computer Quantum Chemistry*, UCH Publishers, Inc. Lybraud T. P., Chap. 8 Computer Simulation of Biomolecular Systems Using Molecular Dynamics and Free Energy Perturbation Methods and Chap. 7 Mezey P.G., *Molecular Surfaces*.
- [Mathews] Mathews C. K. , van Hold K. E. , *Biochemistry*, The Benjamin/Lummings Publishing Company, 1990
- [McCammon] McCammon J. A. , *Dynamics of proteins and nucleic acids*, Cambridge university press, 1987.
- [McQuarrie] McQuarrie, *Statistical Mechanics*, Harper and Row, NY 1976.
- [Offenhardt] Offenhardt P., *Atomic and Molecular orbital theory*, McGraw-Hill, 1970
- [Olson] Olson, W. Topics in Nucleic Acid Structure; Neidle, S., Eds., MacMillan, Ltd. London 1982.
- [Pauling] Pauling, *The nature of chemical bond*, Cornell univ. press.
- [Peña] Peña Luis de la, *Introducción a la mecánica cuántica*, Ed. científico universitarias. UNAM-FCE, México 1991.
- [Richards] Richards E. G. , *An introduction to physical properties of large molecules in solution*, Cambridge university press, 1980.
- [Saenger] Saenger W. , *Principles of nucleic acid structure*, Springer-Verlag, 1984.
- [Szabo] Szabo A. & Ostlund N. S. , *Modern Quantum Chemistry*, Maryland, 1989.
- [Weissbluth] Weissbluth M., *Atomos and Molecules*, Academic Press, New York, 1978.
- [Altona, et al, 1974-94] C. Altona and M. Sundaralingam, Conformational analysis of the sugar ring in nucleosides and nucleotides. A new description using the concept of pseudorotation, *J. Am. Chem. Soc.* 1974, 94, 8205-8212.

- [Arnott, et al, 1974-76] S. Arnott & E. Selsing, Structures of the polynucleotide complexes poly(dA)·poly(dT) and poly(dT)·poly(dA)·poly(dT), *J. Mol. Biol.*, vol. 88, p509, 1974; and S. Arnott, P. J. Bond, E. Selsing and P. J. C. Smith, Models of triple-stranded poly nucleotides with optimised stereochemistry, *Nucleic Acids Res.* vol. 3, 2459, 1976.
- [Bayly, et al, 1993-97] C. I. Bayly, P. Cieplak, W. D. Cornell and P. A. Kollman, A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: The RESP Model, *J. Phys. Chem.*, 1993, 97, 10269-10280.
- [Caldwell, et al, 1990] James Caldwell, Liem X. Dang, Peter A. Kollman, Implementation of Nonadditive Intermolecular Potentials by Use of Molecular Dynamics: Development of Water-Water Potential and Water-Ion Cluster Interactions. *J. Am. Chem. Soc.*, 1990, 112, 9144-9147.
- [Cohen, et al, 1994] Cohen J. S. and Hogan M. E., The New Genetic Medicines, *Scientific American*, December 1994, 50-55.
- [Dang, et al, 1991] Liem X. Dang, Julia E. Rice, James Caldwell and Peter A. Kollman. Ion Solvation in Polarizable Water: Molecular Dynamics Simulations. *J. Am. Chem. Soc.* 1991, 113, 2481-2486.
- [Dervan, et al, 1989] Thomas J. Povsic, and Peter B. Dervan, Triple helix formation by oligonucleotides on DNA extended to the physiological pH range., *J. Am. Chem. Soc.*, 1989, 111, 3059-3061.
- [Falk, et al, 1970] Falk M., Poole A. G. *Canad. J. Chem.* 48, 1536, 1970.
- [Felsenfeld, et al, 1957] G. Felsenfeld, D. R. Davies and A. Rich, Studies on the formation of two and three-stranded polyribonucleotides. *Biochem Biophys Acta* 1957, 26, 457-468, *J. Am. Chem. Soc.* vol. 79, 2023, 1957.
- [Forester, et al, 1991] Forester T. R., McDonald, Molecular dynamics studies of the behavior of water molecules and small ions in concentrated solutions of polymeric B-DNA. *Molecular Physics*, 1991, vol. 72, No. 3, 643-660.

- [Gunsteren,et al,1992] Gunsteren W. F. van and Mark A. E. ,*Review*, On the interpretation of biochemical data by molecular dynamics computer simulation, *Eu. J. Biochem*, 204, 947-961,1992.
- [Hagler,et al,1979] Hagler A.T. & Moult J., *Nature*, Vol.272 (1978),
- [Laughton,et al,1992] Laughton C.A. and Neidle S., Molecular Dynamics Simulation of the DNA Triplex d(TC)5d(GA)5d(CT)5. *J.Mol.Biol.*, 223, 519-529, 1992.
- [Lyubartsev,et al,1995] Lyubartsev A.P.,*J. Phys. Chem.*, 99, 1995.
- [Mergny,et al,1991] Mergny Jean-Louis , Sequence Specificity in Triple-Helix Formation: Experimental and Theoretical Studies of the Effect of Mismatches on Triplex Stability, *Biochemistry* 1991, 30, 9791-9798.
- [Narten,et al,1982] A. H. Narten, W. E. Thiessen and L. Blum, Atom pair distribution functions of liquid water at 25°C from Neutron diffraction, *Science*, 10-sep-1982, 217, 1033-1034.
- [Orozco,et al,1990] Orozco M., Luque F.J., On the Use of AM1 and MNDO Wave Functions to Compute Accurate Electrostatic Charges, *J. Comput. Chem.*, Vol. 11, No. *, 909-923 (1990).
- [Parrinello,et al,1985] R.Car & M. Parrinello, *Phys. Rev. Lett*, vol.55,2471, (1985).
- [Postrino,et al,1994] Postrino P., Ricci M, Soper A. Water above its boiling point: Study of the temperature and density dependence of the partial pair correlation functions. Neutron diffraction experiment. *J. Chem. Phys.*, 101 (5), 1-sep-94.
- [Radhakrishnan,et al,1991] Ishwar Radhakrishnan, Carlos de los Santos and Dinshaw J. Patel., Nuclear Magnetic Resonance structural studies of intramolecular purine-purine-Pyrimidine DNA triplexes in solution, *J. Mol. Biol.*, 1991,221,1403-1418.
- [Singh,et al,1984] Singh U.C. and Kollman P.A., An Approach to Computing electrostatic Charges for Molecules, *J. Comput. Chem.* vol.5 n.2 (1984).

- [Soper,et al,1986] Soper A.K. and Phillips M.G., A new determination of the structure of water at 25 C *Chemical Physics*, 107 (1986) 47-60.
- [Soper,et al,1994] Soper A. K., Orientational correlation function for molecular liquids: The case of liquid water *J.Chem.Phys.*, 101 (8), 15 Oct. (1994).
- [Stein,et al,1993] Stein C.A. and Cheng, Y.C., Antisense Oligonucleotide as therapeutic Agents-Is the bullet really magical?, *Science*, Vol 261, Agosto, 1993.
- [Sun,et al,1993] Jian-Sheng Sun and Claude Hélène, Oligonucleotide-directed triple-helix formation, *Current Opinion in structural Biology*, 1993, 3, 345-356.
- [Swaminathan,et al,1991] Swaminathan S., Ravishanker G., and Beveridge D.L., Molecular Dynamics of B-DNA Including Water and Counterions: A 140-ps Trajectory for D(CGCGAATTCGCG) Based on the GROMOS Force Field, *J.Am.Chem.Soc.* 1991, 113, 5027-5040.
- [Trokhymchuck,et al,1995] Trokhymchuck A.D., Holovko M.F. and Heinzinger K. Density and charge correlations in water, *Mol. Phys* 1995, vol.86, No.4, 797-808; Trokhymchuck A.D., Holovko M.F. and Heinzinger K. On the structure factor for water at small Wavenumbers, *Z. Naturforsch* 50,18-20 (1995).
- [Watson-Crick,et al,1953] J.D.Watson and F.H.C.Crick, 1953a, A structure for desoxyribose nucleic acids. *Nature*, Lond. 171:737-738. 1953b, Genetical implications of the structure of desoxyribose nucleic acid. *Nature acid. Nature*, Lond.
- [Weiner,et al,1984] Weiner S. J., Kollman P. A., A New Force Field for Molecular Mechanical Simulation of Nucleic Acids and Proteins, *J. AM. Chem. Soc.* 1984, 106, 765-784.
- [Weiner,et al,1986] Weiner S. J., Kollman P. A., An All Atom Force Field for Simulations of Proteins and Nucleic Acids, *Jour. Compu. Chem.*, Vol. 7, No. 2, 230-252, 1986.
- [Yang,et al,1995] Yang L, Weerasinghe, S et. al., Dielectric Response of Triplex DNA in Ionic Solution from Simulations, *Biophysical Journal*, Vol. 69, Oct. 95, 1519-1527.

- [Discover] *Discover User Guide*, version 2.96 Biosym Technologies, Inc. San Diego, CA, 1994.
- [Fortran77] *Fortran77*, Notas curso DGSCA-UNAM, Plan de becarios de supercomputo 92.
- [Gaussian 92] *Gaussian 92*, Programmer's Reference and Users Manual, Pittsburgh, PA, 1993.
- [Insight] Paquete *Insight 2.3*, módulos *Decipher*, *Analysis*, *Builder y Biopolymer* User's Guide, Biosym Technologies, Inc. San Diego, CA, 1994.
- [Unichem] *Unichem User's Guide*, Cray Research, Inc, Distrib. center Mendota Heights MN, 1995.